VILNIUS UNIVERSITY

Mindaugas Gilaitis

THE POSSIBILITIES OF SEMANTIC NATURALISM: THE STANDARD VIEW AND AN ABILITY-BASED ALTERNATIVE

Doctoral dissertation

Humanities, Philosophy (01 H)

Vilnius, 2017

The dissertation was prepared at Vilnius University during 2011-2017

Research supervisor:

prof. dr. Marius Povilas Šaulauskas (Vilnius University, Humanities, Philosophy – 01 H)

Research consultant:

prof. dr. Nenad Miščević (Central European University, Humanities, Philosophy – 01 H)

VILNIAUS UNIVERSITETAS

Mindaugas Gilaitis

SEMANTINIO NATŪRALIZMO GALIMYBĖS: STANDARTINĖ KONCEPCIJA IR GEBĖJIMAIS GRĮSTA ALTERNATYVA

Daktaro disertacija

Humanitariniai mokslai, filosofija (01 H)

Vilnius, 2017

Disertacija rengta 2011-2017 metais Vilniaus universitete

Mokslinis vadovas:

prof. dr. Marius Povilas Šaulauskas (Vilniaus universitetas, humanitariniai mokslai, filosofija – 01 H)

Mokslinis konsultantas:

prof. dr. Nenad Miščević (Centrinės Europos universitetas, humanitariniai mokslai, filosofija – 01 H)

Contents

Introduction

Contemporary analytic philosophy is predominantly "naturalistic": the majority of philosophers that belong to the analytic tradition claim to be "naturalists" and the theories they propose are maintained to be developed in the spirit of "naturalism". But the term "naturalism", when used in the context of contemporary analytic philosophy, typically has a specific meaning: primarily, the notion is not used to indicate that the proposed theory is – or a philosophical theory in general ought to be – compatible with the truths of natural or other empirical sciences, but rather to indicate that the theory is a piece of "naturalization" of some allegedly non-natural concepts or the phenomena that they signify (for example, norms, knowledge, action, or colors). From the naturalistic perspective, the categories or entities that are postulated by the natural sciences, such as physics, chemistry or biology, are not problematic. However, according to naturalism, discourse about norms, psychological states of agents or intentional actions are problematic and therefore require naturalization of the relevant categories.

Philosophical naturalism comes in a variety of forms, and the concept of naturalism has no sharp boundaries. But it is safe to maintain that currently it is "scientific naturalism" that is the predominant form of philosophical naturalism in contemporary analytic philosophy. At this point it should be emphasized that what is at issue here is not respect for the results of the natural sciences, or empirical sciences more generally. Rather, the scientific form of naturalism involves the stronger claim that the natural sciences are, or ought to be, the only guide in matters of semantics, epistemology and ontology. In short, scientific naturalism is guided by idea that the description of nature or reality provided by the natural sciences is the only genuine or unproblematic description of it there is. Consequently, there is no genuine or unproblematic knowledge outside natural science (epistemology), to be part of reality is to belong to the ontology of some natural science (ontology), and what is real or

natural is what can be described by the concepts of some natural science (semantics).

So-called naturalization projects or naturalization proposals have been initiated on the basis of the ideals of scientific naturalism. Hence, the idea of naturalization is typically understood as a substantial intellectual attempt to demonstrate that the allegedly non-natural concepts or the phenomena that they stand for can be understood and explained by relying only on the vocabulary (or ontological domain) of some natural science. More specifically, the concept of naturalization implies the attempt to clarify or reform (which verb seems apt depends on one's point of view) the non-natural concepts in terms of the concepts that are used in the natural sciences, to reinterpret them as playing a nonfactual role, or to eliminate them (at least from the vocabulary that is used to describe reality).

In the philosophy of mind, broadly understood, the idea of naturalization that is rooted in the principles of scientific naturalism transforms into the tenet that such mental phenomena as thinking of something, using language in a meaningful way, perceiving an object, or acting voluntarily can be understood and explained only in terms drawn from the natural sciences. Naturalistic theories of intentionality that are the primary object of this thesis are a special case of scientific naturalism. The kind of theories at issue (such as informational semantics, teleosemantics or biosemantics) are commonly grouped under the heading of "semantic naturalism", for despite their differences, all of them argue that intentional and semantic phenomena are "natural" in the sense implied by scientific naturalism. As one of the leading semantic naturalists Jerry Fodor has famously put it (1987), if intentionality is real, it must be something else: there is no place for intentional categories in a physicalistic worldview.

The dissertation is dedicated to the inquiry of the possibilities of semantic naturalism: to the critical analysis of its standard form as well as to the quest and defense of an alternative approach. However, the domain of inquiry of the thesis should be specified more precisely at least in two respects.

First of all, in the analytical and the critical parts of the dissertation, the primary focus of the discussion is on *propositional* form of *intentionality of thought* (in the broad sense of this term, encompassing belief, desire, intention and other related intentional attitudes). This mirrors the focus of semantic naturalism itself, for despite the fact that thought and perception are closely related, and there is good reason to believe that there are non-propositional forms of intentionality, the kind of mental states that have been the focus of naturalization proposals are propositional attitudes where, and the debates primarily revolve around intentionality of thought, broadly construed.

Secondly, the analytical sections of the thesis are dedicated to the analysis of *the methodological and epistemological assumptions* of the orthodox form of semantic naturalism (as well as to the theory of intentionality more generally), and to the analysis of the *presupposed philosophical theories* that are implicit in the predominant paradigm. Thus, the object of research is not some specific naturalistic theory of intentionality (or "psychosemantics", as some naturalists like to call it), but the general framework, which is implicit in the standard version of semantic naturalism, and which informs and guides the specific theories and their development.

Objectives of the thesis. The thesis has the following aims:

1. Clarify the subject matter of a theory of intentionality in general, as well as of its naturalistic form in particular, and propose a comprehensive analysis of its methodological and epistemological assumptions supplemented by argued interpretations of the main categories.

2. Explicate and analyze the philosophical theories that constitute the underlying framework of the predominant form of semantic naturalism, specify their interrelations, and propose a conception of semantic naturalism based on the interpretation of the explicated premises that reflects the general character of the standard view.

3. Critically evaluate the presupposed theories and defend an ability-based approach in the philosophy of intentionality and the ontology of mind more generally.

4. In the context of the analysis of the problem of content epiphenomenalism, reveal the connection between a theory of intentionality or semantics on the one hand, and theory of action on the other, as well as to propose a solution to the indicated problem.

Claims of the thesis. In general, the thesis argues for the ability-based approach to the problem of intentionality by maintaining that it can address the problems faced by the standard theories and is naturalistic enough despite not meeting the reductive standards of the scientific form of naturalism. More particularly, the thesis argues for the following claims:

1. The problem of intentionality (i) revolves around three groups of questions that are distinguished by reference to their subject matter (intentional states, their content and content determining factors) and are characterized by a set of questions concerning the indicated object of inquiry, and (ii) calls for a constitutive account specifying the essence or nature of intentionality. The standard form of semantic naturalism is additionally committed to the idea that the relevant account ought to be reductive and construed in terms of some natural science.

2. A constitutive theory, aiming at the definition of the essence or nature of some phenomenon, ought to be interpreted on conceptualist lines: essence is conceptual, as is the modality implicit in constitutive claims, whereas the requirement that the latter ought to be substantiated by analyses should be taken to call for substantiation by conceptual analysis, which might take a variety of forms (i.e. not necessarily the specification of necessary and sufficient conditions) and be intertwined with empirical inquiry.

3. A theory of intentionality that is based on the categories of *ability* and *power* is a preferable alternative to the predominant naturalistic accounts of intentionality, which are essentially Cartesian and bifurcationist and are based on the categories of *internal states*, *mental representations* and *causal relations* and/or *biological functions* determining their semantic properties. The ability view can (i) explain the problems that the standard naturalistic theories face and have difficulty accounting for, (ii) is compatible with a

version of semantic externalism, which does not imply that external factors unknown to the subject directly determine the intentional content and represents the external environment as a relevant contextual element for determination of content, (iii) can be construed on the basis of the Fregean *Sinn* without commitment to semantic internalism, and (iv) despite the fact that it does not meet the reductive requirements of scientific naturalism is naturalistic enough.

4. The problem of content epiphenomenalism, which is a daunting problem for naturalistic theories of mind and action, can be accounted for if a theory of action that is based on the idea of agent causation, the Aristotelian theory of causation and ability or a power-based ontology of mind replaces the standard view of action, which reduces agency to a concatenation of events (and agent causation to event causation), rests on the Humean theory of causation and a particularist ontology of mind that is implied by the doctrine of token physicalism.

Methodology of the research. The thesis presupposes the doctrine of intentional realism, according to which intentional states and their semantic properties are real, i.e. are part of reality. Intentional realism is assumed not only because it is the near-consensus among philosophers of mind, but also because it is presumed by semantic naturalism itself: as it is shown in section 3.2. of chapter 3 in part I, intentional eliminativism or antirealism can be a view that follows from a failure of naturalization, and not a position that is its starting point.

Furthermore, a distinction between constitutive and causal-enabling accounts or explanations, which is presented and argued for in section 2.4. of chapter 2 in part I, plays an important methodological role in the thesis. Only in the context of this distinction can one comprehend how the ability view can explain the nature of some mental phenomena without proposing (and without being committed to) a causal-enabling explanation of it. Besides, the idea of a bifurcationist conception of mind that is discussed in chapter 2 of part II also depends on a grasp of the idea of constitutive account: bifurcationism can only

be a feature of a constitutive and not of a causal-enabling account of some phenomenon. In other words, something can be bifurcated only from the perspective of its constitutive and not of its causal-enabling story.

The dissertation also uses the methods of logical and conceptual analysis that are necessary for the critical analysis of semantic naturalism as well as for the defense of an alternative view.

Relevance and novelty of the research. Questions about the nature of intentionality and semantics have been at the center of philosophical inquiry not only in the analytic philosophy, but also in the phenomenological and hermeneutic traditions. What it is to think something or think about something? How is it possible to think about something that does not exist or believe something that is false? What it is to understand what is said and what is required for one to say something in a way that others can understand it? What it is that we think, believe or know, say or express in some other way? These and similar questions have been a part of the Western philosophical tradition since its dawn, and not only are they not specific to any one philosophical school or tradition – they can hardly be squeezed into some narrowly defined area of academic research. It takes little effort to reflect on the fact that what one *thinks* or *says* can also be what one *knows*, and what one knows can not only be the basis of how and why one *acts*, but also the essential factor for moral, legal or aesthetic evaluation. Thus, the problem of intentionality and its place in nature is relevant not only to inquiries into the nature of thought, broadly construed, but also to an understanding of knowledge, action and related concepts.

In the analytic tradition, the problem of intentionality has been discussed in a number of articles published in such academic journals as *The Journal of Philosophy, The Philosophical Quarterly, The Philosophical Review, Proceedings of the Aristotelian Society, Mind or Mind and Language,* among others. The nature of intentionality or semantics has been discussed by such prominent authors as Jerry Fodor (1975, 1987, 1990, 1994, 1998, 2003), Hans-

Johann Glock (2001, *forthcoming[1]*), Tyler Burge (1979, 1986, 1988, 2007), Fred Dretske (1981), Ruth Millikan (1984, 1995), Robert Stalnaker (1984, 1999, 2008), John McDowell (1996, 1998a, 1998b), Peter Hacker (2007, 2013), David Papineau (1987), Christopher Peacocke (1992), Donald Davidson (2001, 2004), Hilary Putnam (1975b, 1981, 1988), Stephen Schiffer (1972, 2005), John Searle (1983), Scott Soames (2010), Colin McGinn (1983, 1989, 2002), Tim Crane (2013), Anthony Kenny (1989), among many others. So the abundance of academic interest to the questions that are discussed in the dissertation is manifest.

The relevance and novelty of the present thesis lies in several aspects. First of all, a novel feature of the research consists in its analysis (carried out in the first part of the dissertation) of the methodological and epistemological assumptions of a theory of intentionality in general, as well as of its predominant naturalistic form. The research is also novel in its analysis (carried out in the second part of the thesis) of the general framework of the predominant form of semantic naturalism. The analysis carried out in both of these parts might be relevant not only to the development of current naturalistic theories of semantics. The logical-conceptual analysis of the fundamental assumptions implicit in the orthodox framework might be relevant when aiming for a more radical position, and willing to propose an alternative whose foundations would be based on different categories.

Some novel aspects of the thesis are also present in the third – critical – part of the thesis, where some novel arguments against the representational theory of concepts are proposed, and some new proposals are suggested in the context of a discussion of the implications and theoretical virtues of the defended alternative (for example, with regards to the question whether the ability view is compatible with semantic externalism or, given the current trends in the analytic philosophy – the very important question of whether the

---

[1] I am very grateful to Hans-Johann Glock for having the privilege to get acquainted with the manuscript of his forthcoming book on animal minds.

ability view is naturalistic). Finally, a novel solution to the problem of content epiphenomenalism is proposed at the end of the thesis.

Previous research on the topic. To the best of my knowledge, there is no similar type of academic research on the topic of semantic naturalism. More specifically, there is no research the objective of which would be to clarify the explanatory character – the methodological framework – of the general agenda of semantic naturalism, and no philosophical analysis whose primary goal would be to analyze and reflect on the general framework underlying the currently predominant version of semantic naturalism, which informs the content and guides the development of the specific theories falling under its heading. Having said that, however, Jeff Speak's (2003) dissertation is worth noticing here, for although it is not on the topic of naturalization of intentionality, it is related to the present thesis in its attempt to keep the methodological premises of a theory of intentionality explicit and encompass all three groups of questions that are implicit in the general formulation of the problem of intentionality.

On a more general level, the questions relating to the naturalization of intentionality or semantics have been discussed in the works of John McDowell (1996, 1998a, 1998b), Jennifer Hornsby (1997), Fred Dretske (1981, 1988, 2000), Christopher Peacocke (1992) and Donald Davidson (2001, 2004), among others. In these cases, the problem of intentionality overlaps with questions in the domain of epistemology and philosophy of action (or theory of rationality more generally), thus relating the question of the possibility of naturalization of intentionality to the question concerning the possibility of naturalistic accounts of knowledge and action.

No academic research on the theme of semantic naturalism has been done in Lithuania. However, research on other topics in analytic philosophy of language, philosophy of mind or philosophy of action has been done by Jonas Dagys (2006a, 2006b, 2006c, 2007, 2008, 2012), Rolandas Pavilionis (1981, 2005), Algirdas Degutis (1984, 2007), Mindaugas Japertas (1998, 2001, 2005, 2007, 2012) and Vilius Dranseika (2010, 2012).

Structure of the thesis. The thesis consists of an introduction, three parts, and conclusions.

The research in the first part of the thesis is guided by the question "What is a naturalistic theory of intentionality?". Given that a theory of intentionality by many is considered to be closely related to, or even the same as, a theory of content, this calls for a clarification of the relation between the notion of intentionality and the idea of content (Chapter 1). An explication of what is meant by "the problem of intentionality", as well as an analysis of what kind of explanation or theory is being called for by the problem follow in the subsequent chapter (Chapter 2). The subject matter of the problem of intentionality is being clarified first (Chapter 2, section 2.1.). Then, the character of the kind of explanation or a theory that is being called for – a constitutive one – is discussed in detail, including the discussion of the nature of modality involved in constitutive claims (section 2.2.), the relation between the constitutive claims and the idea of conceptual analysis (section 2.3.), and the distinction between a constitutive theory and theories that are guided by other explanatory interests (section 2.4.). The last chapter of the first part of the thesis is dedicated to the clarification of the notion of "naturalization of intentionality", which consists of a discussion of the idea of naturalism (section 3.1.), a clarification of the relation between the agenda of naturalization and the doctrine of intentional antirealism, and a conclusive part as to what "naturalization" of semantics or intentionality amounts to (section 3.2.).

The second part of the thesis is dedicated to a piece of philosophical analysis of the general framework underlying the standard version of semantic naturalism. The first sketch of the general picture is made by explicating the fact that the predominant naturalistic theories of intentionality belong to a group of causal theories of mental content, and a more general cluster of causal theories of mind (Chapter 1). The framework of causal theories of mental content then becomes the basis for identifying some of the central problems that naturalistic theories of semantics face and are having troubles to account for, as well as for an introduction of the idea of the bifurcationist picture of the

mind which appears to be manifest in causal theories of mind and implicit in the analysis of intentional attitudes implied by the framework of semantic naturalism (Chapter 2). The roots of the bifurcationist analysis of intentional attitudes are then examined: first, the externalist character (Chapter 3, section 3.1.) of naturalistic theories of intentionality, as well as the idea of semantic externalism more generally, is clarified; then, the notion of the Cartesian target is discussed (section 3.2.). The successive sections are devoted to the analysis of the origins (Chapter 4, section 4.1.) and interpretation of functionalism (section 4.2.), as well as the discussion of the representational theory of mind and concepts (Chapter 5).

The third part of the thesis consists of two chapters. The first chapter is dedicated to a critical discussion of the representationalist framework and a defense of an alternative ability-based approach to the philosophy of intentionality and the ontology of mind more generally. First, the idea that concepts are mental representations is being examined and criticized (Chapter 1, section 1.1.). Then, an alternative approach to the mind that is based on the categories of ability and power is being introduced and its general framework explained (section 1.2.), which is succeeded by the application of the ability view to the question of the nature of concept possession and concepts (section 1.3.). This, in turn, leads to a more general evaluation of the ability-based approach which amounts to a discussion of its theoretical virtues and implications with regards to specific problems and philosophical positions (section 1.4.). The first chapter ends with a critical evaluation of the so-called descriptivist account of the Fregean *Sinn* (section 1.5.) showing that, contrary to the prevalent opinion, the idea of Fregean *Sinn* does not imply semantic internalism.

The second chapter of the third part is dedicated to a detailed analysis of, and a proposal of a solution to the problem of content epiphenomenalism. The chapter consist of an explication of the problem, the discussion of the character of intentional explanations (Chapter 2, section 2.1.), the idea of mental causation and the standard view of action (section 2.2.), and ends with a

defense of a conception of agency that is based on the idea of agent causation, Aristotelian theory of causation, and ability or power-based ontology of mind (section 2.3.). If accepted, the alternative, processual and agent-based account of action can show the way out of the problem of content epiphenomenalism thus indicating the way action theory is related to a theory of intentionality.

I. What is a naturalistic theory of intentionality?

1. Intentionality and content

"Intentionality" is a philosophical term that is used to characterize a feature of our world, the main character of which is commonly described in terms of a capacity of something – a thought, a sentence, a photograph or a gesture –to *be about* something else, to *stand for* something, or to *present* something as being some particular way. Natural languages contain several terms that historically have been used to pick out this phenomenon. Some verbs that are used to that end include "mean", "refer", "denote" and "signify", and their nominalizations – "meaning", "reference", "denotation" and "signification" – pick out the same phenomenon as a technical notion of "intentionality".

The etymological roots of this mediaeval term, which was reintroduced in the XIX century by Franz Brentano in his *Psychology from an Empirical Standpoint* (1874), provide the basis for a metaphorical description of this evidently perplexing capacity. The term derives from the Latin verb *intentio* which is derivative from *tendere* meaning *stretching* or *extending*, and it is by reference to these ideas that the phenomenon is usually characterized: *being intentional* is *being directed* towards something or some end.

In the first place, the topic of intentionality in the philosophical tradition has been primarily discussed within philosophy of mind, both in analytic and phenomenological tradition. Brentano not only has defined intentionality as the directedness of the mind towards its "content" or "object" (Brentano 1995: 68); contrary to then prevailing Cartesian tradition, he too has maintained that intentionality, and not consciousness, is the essential and distinguishing feature of mentality: in Brentano's view, only what is intentional can be mental.

Brentano's intentionalistic conception of mentality does not have many adherents these days, but independently of what one holds to be *the* distinguishing feature of mental phenomena, if there is one, everyone but

17

eliminativists agrees that at least *some* mental phenomena not only have this feature, but are partly defined by reference to it. For example, pains, tickles or moods by common (yet not universal) agreement are not intentional states whereas thoughts, beliefs, desires, wishes and intentions are.

But the scope of the topic of intentionality should not be restricted only to the domain of philosophy of mind or philosophical psychology. For the ideas of being about something else or presenting something as being some particular way are also implicit in the ideas of meaning, referring, signifying, and their cognates. Thus, intentionality constitutes the core of philosophical theories of meaning and reference, and plays a pivotal role in philosophy of language and philosophical semantics. In the analytic tradition, Wittgenstein's picture theory of meaning (Wittgenstein 1922), Russell's theory of judgement (Russell 1905, 1966) and Frege's theory of *Sinn* and *Bedeutung* (Frege 1993) are cases in point.

One can distinguish several forms of intentionality on the basis of the grammatical structure of intentional attitude reports. John might not only *believe* or *known that the Earth rotates around the sun*; he might also *wish to go to a museum* or *think about the Pythagorean theorem*. As Hans-Johann Glock notes (2001: 107), intentional verbs can occur in three sentential forms:

(1)   A   Vs (thinks/believes/hopes, etc.)   that *p*

(2)   A   Vs (intends/wishes/plans to, etc.)   to ϕ

(3)   A   Vs (loves/desires/thinks about, etc.)   *X*

All three forms reflect the abstract and most general structure of intentional attitude reports, where "the verbs that can replace 'V' denote different types of *intentional attitudes*, 'A' their *subject*, and 'that *p*', 'to ϕ' or '*X*' their *content*" (ibid.).

It is common ground among philosophers working on the topic of intentionality that sentences of the form exhibited in (1) express what are commonly known as *propositional attitudes* (i.e. their content is a

proposition[2]), sentences of type (2) – *action-oriented attitudes* (i.e. their content is action), and sentences of type (3) are taken to express *object-oriented attitudes* (i.e. their content is an object of some kind, though not necessarily an existing one). Despite these differences in their contents, all three forms of intentionality can exhibit what Brentano took to be one of its defining features – namely, intentional inexistence. One can believe or think something that is not the case (or believe or think a falsehood), intend to do something which does not happen or expect someone who does not exist. However, not all mental states that are directed towards something can exhibit intentional inexistence, and factive states like *knowing*, *seeing* or *being aware of* evidently cannot. So, if states signified by factive verbs are intentional, then contrary to Brentano, the idea of intentionality does not imply the idea of intentional inexistence, and it is better to think of it in terms of *aboutness* or *directedness*.

In the analytic tradition, the topic of intentionality is commonly approached via the idea of content, and a "theory of intentionality" is commonly used interchangeably with a "theory of content". A possible explanation of this can consist in the ways the notion of "content" is commonly used in these debates – the way indicated in the above explications of the three sentential forms of intentional attitude reports. From the above descriptions of different forms of intentionality, it is clear that the notion of "content" applies to that, possession of which is sufficient for the bearer of that content to be

---

[2] The term "propositional attitude" might be read in two different ways. The term was originally introduced by Bertrand Russell (1912) to express the idea that believing, hoping, thinking (etc.) that something is so *is a relation to a proposition*. However, more recently it has also been used as term, under which believing (etc.) that something is so is subsumed without prejudice to what it is. In other words, without taking a stance on whether V*ing* that *p* is a *relation* to a proposition. In this non-committal sense, a propositional attitude is simply what a psychological verb followed by a that-clause signifies, whatever that turns out to be. Furthermore, in the philosophical literature the term "proposition" is generally used as a *technical* term to refer to *what* A V*s*, whatever it turns out to be: a structured abstract object composed of an object, a property and a relation (a Russellian proposition) or composed of Fregean *Sinne* (a Fregean proposition), or an unstructured abstract object as in possible world semantics. One might disagree as to how fine-grained they must be, whether they are pleonastic entities (as in Schiffer 2005) or cognitive (as in Soames 2010), or on other issues. But if "proposition" is understood in a technical sense, then it cannot be the case that A V*s* that *p* and yet *what* A V*s* is not a proposition. It would be like saying that V*ing* that *p* is not a propositional attitude in the non-committal sense.

about something. This is trivial in forms (2) and (3), where the idea of content applies to what the state is about: *Marry wishes to leave the black and white room* (content = the black and white room, and the wish is *about* the black and white room), *John thinks about the Chinese room* (content = Chinese room, the thought is *about* the Chinese room). However, in cases of form (1), there is a distinction to be made between the content of a belief – e.g. *that the moon is full* – and what that content is about – namely, the moon. This explains why, when interpreting intentional attitude reports of form (1), the notion of "content" is commonly applied to *what* is thought, and not to what is being thought *about*. However, even in such cases having content is a sufficient condition for being about something else, yet it has that content because it is about that something. So, it is inconceivable that, say, a natural language sentence, when used, expresses some content (i.e. something is being said by the use of a sentence, say, *that Wittgenstein never met Frege after 1914*) and yet it is *not* about what the subject-term refers to (i.e. that it says *that Pegasus has wings* and it is *not* about Pegasus). The same holds for mental states: if a belief has a content *p* (say, *that John likes sailing*) or an instance of perceptual experience has a content *q* (say, *that John is sailing*), then it is sufficient that both the belief and the perception are directed towards their "object".

Usually (1) is regarded as a basic form of intentionality and it is commonly held that (2) and (3) can be reduced to it. This would imply that *all* intentionality (or having of content) is propositional. Surely, the actual content of the implication depends on what being propositional amounts to – a highly controversial matter. But irrespective of which account of propositional content is the correct one, the idea that action-oriented or object-oriented attitudes are propositional (or that they have propositional form) is contentious. What might be a less controversial idea is that the non-propositional forms of intentionality are connected with the propositional ones, and that having the latter might be necessary for being able to have at least some of the former. So even if it is unlikely that when *Peter is thinking of Zeus*, his content – in this case, Zeus – is a proposition, the idea that in order for John to be able to think of Zeus, he

must hold at least *some* propositional attitudes is more plausible. In fact, it seems very likely that if, say, *John wishes to go to a museum*, it is necessary that he at least believes *that museum exists*. But the relation between the propositional and non-propositional forms of intentionality is not direct, and there might be non-propositional forms of intentionality that do not depend on having attitudes of propositional form. One way or the other, there is no good reason to think that (1) is the most fundamental form of intentionality, less so that other forms can be reduced to it.

2. The problem of intentionality and its methodological aspects

2. 1. What the problem is about

As Peter Hacker points out, philosophical reflections on intentionality go back to the "Parmenidean and post-Parmenidean reflections on the possibility of thinking what is not the case, or thinking of what does not exist" (Hacker 2013: 60): Plato expressed the same puzzlement by raising the question of how it is possible to think "something that is not" (Thaetetus 189a) while Wittgenstein centuries later wondered "How can one think what is not the case?" (Wittgenstein 1958: 31). In fact, the problem of falsehood or intentional inexistence is considered by some as *the* central key to the perplexing character of intentionality. For example, Tim Crane holds the view that "unless we understand non-existence we cannot understand intentionality", meaning not the nature of existence, but the "phenomenon of thought about the non-existent", or rather "what it means to think about the non-existent" (Crane 2013: ix).

It is true that falsehood or intentional inexistence are in important ways related to the phenomenon of intentionality, and that clarification of these concepts might contribute to a better understanding of at least some forms of intentionality. However, neither falsehood nor intentional inexistence has been the central focus of philosophical debates on the nature of intentionality. Of

course, what is and what is not in the focus of a philosophical debate depends on various factors, and some of them might have little, if anything, to do with the philosophical importance of an idea. But there are at least two reasons why the problem of falsehood or intentional inexistence might not the central sun of the debates on the phenomenon of intentionality.

The first reason is the aforementioned fact that not all mental states that are directed towards something can exhibit intentional inexistence, factive states being a case in point. So the problem of intentional inexistence does not arise for factive intentional states. The second reason is the view that the problem of intentional inexistence can be understood by developing a conception of content and its possession that would provide the means to account for the phenomenon that some hold to be the key to understanding intentionality. Hence, it is not the nature of intentional inexistence, or "non-existence", but the nature of content and its possession that are at the center of the philosophical inquiry into the nature of intentionality. Nothing other than intentionality itself can stand at its center.

It is common to speak of the problems surrounding the phenomenon of intentionality in the singular, as if they were a single problem. For example, in Victor Caston's view, "the problem of intentionality is the problem of explaining what it is in general for mental states to have content" (Caston 2007: §2). According to Jeff Speaks, it is "the problem of saying what it is for something – a mental state, an expression of English, a gesture – to represent the world as being some way" (Speaks 2006: 430). Whereas in Erich Ammereller's view, it is the problem of the "intrinsic directedness upon what is the case if I believe truly" (Ammereller 2001: 61) and revolves around questions like "what makes my belief the belief *that p*? What must be the case for a belief to have a certain content? What, for instance, must be the case for my present belief to have the content *that George is coming tonight*, rather than, e.g. *that Sarah is coming tonight*?" (Ibid.).

However, the singular mode of description might be misleading, for it conceals the complexity of the questions (or their clusters) that are implicit in

what is commonly presented as a "general problem". The underlying puzzles can be brought into view by means of the grammatical structure of intentional attitude reports. Glock suggests (2001: 106-107) that we should distinguish between at least three groups of problems on this basis, though they are neither mutually exclusive nor sharply defined.

The first, "mental" group concerns *intentional states* and contains variants of familiar questions concerning the nature of mind, language and their place in nature. According to Glock, "the flair of the mental problems is summed by 'How is it possible for us to have beliefs?'" (Ibid.: 106), but it encompasses questions like "What is it to have a mental state?", "What does being a competent language user amount to?", "How can the mind can make a difference in the natural world?", and more pertinent to topic of this thesis – "Can intentionality be understood as a purely natural phenomenon (in the to-be-specified sense)?".

The second, "logico-semantic" group concerns the *content* of these states and include questions concerning the analysis of intentional verbs, of "that-clauses", and of concepts like *content*, *proposition*, *fact*, etc. The logico-semantic problems can be summed up by the question "What do we think (mean, believe, know, etc.)?", but again, it involves several specific questions like "What is it that we think?", "What is a proposition?", and "Does propositional content have constituents, and is it structured?", among others.

The last group of problems, "evolve around the question: What determines the content of a particular belief (desire, statement, etc.)?" (Ibid.: 107). What, for example, determines that A is thinking that p rather than q, or about x rather than about y? Furthermore, it is a fact that any propositional thought (or a sentence expressing a propositional content) is about its "object": John's belief that Pegasus does not exist is about Pegasus, Mary's thought that apple is red is about apple, and so on. But what determines *that* fact – more generally, the fact that A's V*ing* that a is F is about a? These are the questions that are at the core of what might be called "content-determination" group.

A central mental question that is of great importance for the overall theory of intentionality is the question of the relation between the intentionality of thought and the intentionality of language, and it is commonly presented in terms of explanatory priority: "should linguistic meaning be explained as resulting from mental content, or vice versa, or should the same account be given of both with equal priority" (Thornton 1998: 1). As Tim Thornton rightly points out, within contemporary theories of intentionality,

> the standard approach is to attempt to explain the meaning of sentences as an abstraction from the meaning of utterances made using them. The meaning of utterances is then supposed to be explained as deriving from the content of beliefs or other propositional attitudes that the utterances were intended to convey. […] A different and independent account would then have to be found of how mental states possess their content. (Thornton 1998: 11)

The above strategy is what Speaks calls "the priority of mental content over public language meaning" where "facts about the contents of the mental states of agents are prior to and independent of facts about the meanings of expressions in public languages spoken by those agents" (Speaks 2006: 430). The view that mental content is prior to linguistic meaning, and so the intentionality of language should be explained in terms of the intentionality of thought (or other propositional attitudes) is an idea that is rooted in and has been developed on the basis of intention-based semantics: an approach to foundational semantics influenced by Paul Grice's attempts (Grice 1957, 1969) to explain linguistic meaning in terms of speaker's beliefs and intentions.

This, in turn, leads to an important point about the use of the notion of a "theory of intentionality" or a "naturalistic theory of semantics/content" in the analytic philosophy of mind. The general concept of *a theory of intentionality* does not differentiate between all the different forms of intentionality (A *sees x* vs A *sees* that *p* vs A *thinks* that *p*, A *thinks* of *x*, etc). But in many cases, the terms "naturalistic theory of intentionality" or "naturalistic theory of semantics", or even a general notion of a "theory of intentionality" are used to

apply to paradigm cases of propositional forms of intentionality of thought. This is noted by Barry Loewer in "A Guide to Naturalizing Semantics". First, he rightly points out the predominant view on the priority question:

> Semantic predicates – *is true*, *refers*, *is about*, *has the truth conditional content that p* and so forth – apply to various items, most centrally to natural language expressions and to mental states and events (types and tokens). For example, both the sentence "The cat is crying" and the belief that the cat is crying are about the cat and possess the truth conditional content that the cat is crying. It is generally (and correctly) held that the semantic properties of natural language expressions (and other non-mental representations) are derived from the semantic properties of mental states. (Loewer 1997: 108)

Then, Loewer adds that "the semantic properties of mental states are what makes them intentional states" (ibid.), and he is right on this. Mental states are intentional because they are about *x*, or have a truth conditional content that *p*, etc., and the reason why they are intentional is because this is what "being intentional" *means*. So semantic properties of mental states "make them" intentional in the same way in which a particular animal is "made" into a vixen by being a female fox.

Finally, Loewer points out that "the mental states that have been the focus of naturalization proposals are the propositional attitudes; desire, belief, and perception (perceptual belief)" (ibid.). The list perhaps is not meant to be taken as exhaustive, for it should certainly include thought, desire, and intention, among some other cognitive states that are not necessarily *propositional* attitudes, for the reasons already stated (see Chapter 1). But Loewer is right that the primary focus of what he calls "naturalization proposals" like informational semantics or biosemantics is on *propositional* forms of intentionality, even if not all intentionality is propositional. Furthermore, naturalistic accounts of semantics have primarily focused on the *intentionality of thought*. For example, informational semantics is presented in the following way:

> Informational semantics takes the primary — at least the original —
> home of meaning to be the mind: meaning as the content of thought,
> desire, and intention. The meaning of beliefs, desires, and intentions is
> what it is we believe, desire, and intend. […] So for informational
> semantics the very existence of thought and, thus, the possibility of
> language depends on the capacity of (some) living systems to transform
> information (normally supplied by perception) into meaningful
> (contentful) inner states like thought, intention, and purpose. (Dretske
> 2009a: 381)

So the primary focus of naturalistic theories of intentionality has been the intentionality of *thought* (including belief, desire, intention; hence, "thought" broadly understood) which is assumed to be closely related to the intentionality of *perception*. The latter, it is generally agreed, is the source of information *for* intentionality of thought. In fact, the relation between intentionality of thought and perception seems to be even closer. For example, some hold that both thought and perception have the same kind of content, so what A sees to be the case is the same kind of content as what A believes to be the case, although there is no agreement on whether what is of the same kind is conceptual (as in McDowell 1996) or not (as in Stalnaker 1998a). Neither does there seem to be agreement on whether perceptual content is conceptual (as in McDowell 1996) or not (as in Evans 1982). But irrespective of these differences in opinion, it is generally agreed – and with good reason – that there is a close connection between perception and thought.

To conclude, despite the fact that the critical analysis of the framework of semantic naturalism aims at a general framework that underlies it, the scope of the analytical (Part II) and critical (Part III) parts of this research is focused on a discussion of what is the primary focus of naturalistic theories of intentionality: namely, *propositional forms of intentionality of thought* (i.e. to *what* A thinks/believes/knows/etc.) as opposed to intentionality of perception (i.e. A sees *x* or that *p*), although some discussion of perception (in particular, a

discussion of a causal theory of perception) that is relevant for more general purposes shall be included.

2. 2. Constitutive claims and conceptual modality

According to Speaks, the philosophical problem of intentionality or representation

> calls for an answer which does not merely tell us contingent facts about the way that representation happens to work in our linguistic community, among humans, or even in the actual world; rather, what is sought is an account of the conditions under which, in any possible world, something represents the world as being a certain way. *This is not an arbitrary constraint, but rather is a general feature of philosophical questions about the natures of things*. (Italics – M.G.) (Speaks 2006: 430)

A question about the nature of X is a question about *what it is* to be X, or a question about its *essence*. In that respect the question belongs to ontology – a branch of metaphysics that by general agreement is dealing with two main questions: (1) "What kinds of things exist?" and (2) "What is the nature or essence of these kinds?". The two questions are interrelated, because any ontological claim about *what there is* must not only say *that* something exists (say, that there are leafs or beliefs). It must also be accompanied with an explanation or description of *what* these leafs or beliefs are. But to say *what* they are, and thus to talk sense about them, *is* to specify their nature or essence, and in doing that one gives an answer to the question *what it is* for something to be a leaf or belief, or to be X more generally.

Questions about *what it is* to be or have X are commonly called "constitutive questions". As Tyler Burge puts it,

> a constitutive question concerns conditions on something's being what it is, in the most basic way. Something cannot fail to be what it is, in this

27

way, and be that something. Constitutive conditions are necessary and sufficient conditions for something's being what it is in this basic way. (Burge 2010: vi)

Thus, a philosophical – more specifically, ontological – question about the nature or essence of X calls for a *constitutive* account or explanation of X. Mark Greenberg's understanding reflects a general view. He writes that

by a "constitutive account," I mean the kind of elucidation of the nature of a phenomenon that theorists have tried to give for, for example, knowledge, justice, personal identity, consciousness, convention, heat, and limit (in mathematics). An example is Locke's view that facts about personal identity obtain in virtue of facts about psychological continuity. (Greenberg 2005: 2)

In the passage quoted above, Burge claims that constitutive conditions are necessary and sufficient conditions for something to be X. But one should add here that the features specified in the biconditional should not just *in fact* be possessed by all and only things that are X. It should also be *necessary* that all and only things that are X possess these features for them to be constitutive features of X. In other words, a constitutive account of X must be a necessary truth.

However, as Greenberg notes,

giving a constitutive account is not the same as specifying modally necessary and sufficient conditions. For example, being a type of polygon that can be inscribed in a circle regardless of the lengths of its sides is necessary and sufficient for being a three-sided polygon. But a correct constitutive account of a triangle will plausibly mention the latter but not the former property. (Ibid.: 2)

It is generally agreed that a constitutive account seeks to provide something that cannot be captured in purely modal terms, because the examples like the one with polygon seem to show that there are necessary and sufficient conditions for being X that do not figure in an account of *what it is* for something to be X.

There is no general agreement as to what exactly more is needed, but the general tendency is suggestive. For example, Kit Fine (1994) argues that the notion of "essence" which plays a central role in the metaphysics of identity cannot be captured in purely modal terms, and Paul Boghossian suggests (1989: 532-535) that the additional requirement is that of intensional equivalence. One way or the other, at this point one can conclude that a constitutive account of X is an inquiry into the essential nature of X: of what something *must* be for that something to be X. The question remains, however, as to what is the source or nature of the necessity or modality involved.

Speaks rightly points out that "a natural and traditional starting point is to require that constitutive claims be substantiated by *analyses*" (Speaks 2003: 5), which in his view amounts to showing that "if a class A of facts about meaning or content is constituted by and derived from a class B of such facts, then we must be able to give metaphysically necessary and sufficient conditions for the A-facts obtaining in terms of the B-facts" (ibid.: 6). This view is in fact shared by many theorists working on the nature of intentionality, although often it is an implicit working hypothesis rather than an explicit methodological claim (see e.g. Searle (1983), Fodor (2001: 1-2), Schiffer (1972: 1-6)). However, Speaks' proposal needs clarification, for neither it is obvious what *metaphysical necessity* amounts to, nor is it obvious what an "analysis of X" means, or how it can substantiate the alleged metaphysical truths of constitutive claims.

It is plausible to assume that the reason why the relevant kind of necessity or modality is usually called "metaphysical" is that constitutive accounts of X aim to specify the essential nature *of X*. So the necessity in question seems to be *de re* necessity, and for this reason it is called "metaphysical".

However, the idea of *de re* necessity has been questioned. As Quine has argued, there is no principled way of assigning some properties as essential to a thing and others as accidental (Quine 1980: 148, 155-156), and what properties a thing possesses *essentially* depends on how it is described. In his view,

the doctrine that some of the attributes of a thing (quite independently of the language in which the thing is referred to, if at all) may be essential to the thing, and others accidental […] leads us back into the metaphysical jungle of Aristotelian essentialism. (Quine 1966: 175-176)

Quinean skepticism towards concept-independent *essences* leads to the view that the so-called essential properties are merely properties entailed by some currently salient description. This, in turn, implies that essence is essentially conceptual, as is the modality involved when stating the constitutive conditions of X that must obtain in all possible worlds.

Skepticism towards *de re* necessity naturally leads one to believe that what goes by the name of "metaphysical necessity" is a kind of modality that is essentially conceptual. This position could also explain why, according to some, a constitutive account of the nature of X should amount to intensional equivalence.

However, as Jason Bridges points out,

constitutive claims are ubiquitous inside philosophy and out, and may be defended on grounds of a widely divergent character. At one extreme, a claim about what it is to be an f may be based on considerations that are alleged to emerge simply through armchair reflection on the concept of an f. At the other extreme, and in particular when f's constitute a natural kind, a claim about what it is to be an f may be presented as the empirical deliverance of a natural science. (Bridges 2006: 529)

Bridge's point is not about the kind of modality that is involved in constitutive claims, but about the allegedly different ways to defend constitutive claims. However, the second "extreme" position might be taken to suggest that at least *some* constitutive claims – namely, those involving natural kind terms – cannot be conceptual. For one reason or other, one might hold that the truth of constitutive claims involving natural kinds does not "emerge simply through armchair reflection on the concept" of the relevant natural kind term. And if it does not emerge in this way, then it might seem that at least some constitutive

truths cannot be conceptual. Hence, the idea that what goes by the name of "metaphysical necessity" is conceptual cannot be right.

But the conclusion need not follow, for there is no single interpretation of constitutive claims involving natural kinds terms. So even if one might agree that in such cases, constitutive claims are not the result of the armchair reflection on the relevant concept (whatever that means), that need not preclude a conceptualist interpretation of their necessity. But this is not the standard view.

The predominant account of the so-called theoretical identities involving natural kind terms like "water", "gold" or "tiger" is based on the ideas of Hilary Putnam and Saul Kripke. Their account of theoretical identities is partly based on their semantic theory of natural kind terms: in particular, a causal theory of reference and, in Kripke's case, the idea of rigidity. But neither of these semantic theories is sufficient to explain the necessity of theoretical identities, so ultimately their account of constitutive claims involving natural kind terms rests on the idea of essential property that is concept-independent. Both Kripke and Putnam hold that each natural kind has its own nature (Kripke 1980: 135; Putnam 1975b: 140), its essence (Kripke 1980: 138) or essential properties (ibid.: 133). As Hanoch Ben-Yami points out, "Kripke and Putnam never discussed in sufficient detail how essential properties are determined. However, all the examples of essential properties they mentioned are of *structural* properties" (Ben-Yami 2001: 162). Putnam explicitly says that "if there is a hidden structure, then generally it determines what it is to be a member of the natural kind, not only in the actual world, but in all possible worlds" (Putnam 1975a: 241).

So the general idea of this position is that natural kinds have essential properties that are structural and might be hidden, and constitutive truths about natural kinds that refer to such essential properties (say, "water = H_2O") are true in all possible worlds. Accordingly, constitutive accounts of the nature of natural kinds state truths that are necessary yet a posteriori, and the necessity expressed is considered to be metaphysical, not conceptual. The reason why it

is not a conceptual truth is because it is not part of the concept of water that water is H_2O, and needless to say that it is not the result of the "armchair reflection" on the concept of water or some other natural kind. To sum up, according to the Putnam-Kripke account, the truths of constitutive claims involving natural kind terms are metaphysically necessary, a posteriori, and not conceptual.

But essentialism need not be presupposed in order to account for what goes on in cases of constitutive claims that involve natural kind terms, and the seemingly a posteriori character of such propositions need not imply that their necessity cannot be explained on the conceptualists lines. Accordingly, the idea of metaphysical necessity need not be explanatorily basic.

Alan Sidelle, for example, has argued that the so-called metaphysical necessities discovered a posteriori can be seen as resulting from two parts: the first being a contingent discovery, which is that H_2O is the microstructural composition of water; the second being a convention that whatever microstructural composition paradigmatic samples to which the natural kind term "water" is being used to apply actually have, in any possible world, nothing will count as water unless it has that chemical structure (Sidelle 1989: Ch. 2-3). This view is based on the idea that "modality does not find its home in the mind-independent world, but rather in us, in our ways of speaking and thinking" (ibid.: 1-2).

Sidelle's account of modality is essentially similar to that of Amie Thomasson, who defends a position that she calls "modal conceptualism". In her view, "all modal truths are ultimately based on analytic truths in the sense that modal truths are either analytic truths or based on combining an analytic truth with an empirical truth" (Thomasson 2007: 62-63). Consequently, according to this position, even if the fact that water is H_2O is not part of the concept of water, and even if it is *de re* in the sense that it is about the designated substance, the ultimate source of the necessity of a constitutive claim "water = H_2O" or "(x) x is water if and only if x is H_2O" is conceptual, as is the nature of the modality involved.

What both Sidelle's and Thomasson's views lead to is a conception of modality, according to which "'metaphysical' inquiries into the essence of natural kinds are investigations of complex conceptual connections and their interplay with contingent facts" (Glock 2003: 101). Consequently, as Peter Strawson has put it, "the vogue for what are called metaphysical necessities and described as necessary identities established a posteriori" are to be regarded "as yet another conceptual decision or revision, reasonably enough adopted in the light of empirical discovery" (Strawson 1998: 66). Such a conception of the relevant kind of modality *might* undermine the idea that all constitutive truths emerge simply through armchair reflection on the concept. But it need not result in the conclusion that the so-called metaphysical necessities are not essentially conceptual.

On the other hand, one might wonder what the relation between conceptual truths and empirical inquiry is. Here one might quote Strawson, who described the relation in the following way:

> empirical discovery and natural scientific and technological advance call for conceptual revision, adjustment, or even decision. Banal examples are those of the concept of 'fish', revised to exclude the mammalian whale, and the concept of 'mother of', requiring adjustment to cope with the fact that the bearer of a child may not be identical with the supplier of the fertilized ovum. (Ibid.: 65-66)

So according to Strawson, empirical truths or discoveries might call for conceptual *change*. But one should take note here: it is not that conceptual truths can be empirically discovered, but that empirical truths can call for a reasonable conceptual revision. So there is a relation between the two, but of a kind that does not preclude of there being a distinction between empirical propositions and conceptual ones. Furthermore, if there can be conceptual change, then it is surely incompatible with the view that conceptual truths cannot be revised. For some this might be sufficient to consider conceptual truths as being impure. One way or the other, they are not metaphysical in any real sense.

There might still be an option for an advocate of non-conceptual metaphysical necessity. For one might think that if constitutive truths are to be substantiated by analyses, this cannot amount to their being *conceptual* truths, because the analysis in question is an analysis of the nature *of X*, and not *of the concept of X*. In other words, the analysis is metaphysical, as is the necessity involved, because it is *about X*, and not *about the concept of X*. Consequently, the analysis of X cannot be a piece of *conceptual* analysis.

2. 3. Conceptual analysis and the false opposition

In order to answer the question whether an analysis of X can be *conceptual* analysis, one must first clarify what the notion of "conceptual analysis" means. But instead of doing that directly, it might be more suggestive to begin with a specific example of what is generally considered to be a case of philosophical analysis – namely, the analysis of knowledge.

It is true that in many cases, both written and oral, the analysis of knowledge is presented as it appears to be: as an analysis of *knowledge*, and not as an analysis of *the concept of knowledge*. Many epistemologists, at least until the groundbreaking paper by Edmund Gettier (1963), seemed to agree that the analysis should take the traditional form of a biconditional that is a necessary truth. Assuming the implicit necessity of the biconditional, the result of the analysis should fill the "[…]" part in "A knows that *p* if and only if A […] that *p*". Gettier's paper is generally taken to result in a refutation of what is known as the tripartite analysis of knowledge in terms of justified true belief (JTB). For many epistemologists, this meant that the analysis must be supplemented by an additional "degettierization" clause (+D), and should take the form of "A knows that *p* if and only if A [JTB+D] that *p*".

However, new analyses were susceptible to further counterexamples. For some this indicated that the analysis of knowledge in terms of *belief* cannot be given, which, in turn, for some implied that no analysis of knowledge is

possible. For others, a vast amount of counterexamples showed that the analysis in terms of stating necessary and sufficient conditions cannot be attained, which, again, for some implied that no analysis of knowledge is possible, although for others it meant that it might be possible in some other form.

The above possibilities are not meant to be a general overview of post-Gettier epistemology, but rather a demonstration of the possible *forms* that the *analysis* of X might *take*. Some further cases meant to clarify the notion of an analysis of X will be discussed later on. For now, it should suffice that what has generally taken to be an analysis of knowledge *itself* has *also* been described as a case of *conceptual* analysis of the *concept* of knowledge. But in order to get an idea of what the alternative mode of presenting the inquiry amounts to, one should not only clarify the notion of "conceptual analysis" or "analysis of concept". It might also be suggestive to get a better grip on the notion of "analysis" that has been used in the analytic philosophy since the "linguistic turn".

One kind of analysis that appeared in analytic philosophy after the linguistic turn was *logical analysis*, which aims to uncover the *logical form* of propositions, and takes the form of a *logical* paraphrase of the ones that are philosophically puzzling (e.g. "The king of France is bald"). The primary example of logical analysis is Bertrand Russell's theory of descriptions (Russell 1905) – a "paradigm of philosophy", according to Frank Ramsey (Ramsey 1931: 263).

Another kind of analysis was *conceptual* and *decompositional*, for one of its central aims was defining complex concepts in terms of simpler ones ("decomposing"), up to the point at which one has reached indefinable simple concepts. George Moore can be considered as a proponent of this type of conceptual analysis by means of which he reached the well-known meta-ethical conclusion that *goodness* is one of the indefinable simple concepts (Moore 1903). However, as Glock notes,

> Wittgensteinians and Oxford conceptual analysis reject the idea that propositions have ultimate components or even a definite structure. As a result, analysis in their hand means neither decomposition into ultimate or more basic components nor logical paraphrase. Instead, it means the explanation of concepts and the description of conceptual connections by way of implication, presupposition and exclusion. (Glock 2008: 158)

Strawson's *connective analysis* is an obvious case of Oxford conceptual analysis, although as Strawson himself pointed out (Strawson 1992: 19) the notion of "analysis" might be misleading, and it might be more appropriate to speak of "elucidation" instead. At this point, one could also mention Gilbert Ryle's "occamizing" dictum, which aptly represents the general attitude of Oxford conceptual analysis towards philosophical problems: in Ryle's view, "philosophical problems are problems of a certain sort; they are not problems of an ordinary sort about special entities" (Ryle 1971: vii).

To complicate the picture even more, one could supplement the list with Rudolf Carnap and Willard V. O. Quine, for in their case analysis took a yet another form. As Glock points out,

> for both Carnap and Quine analysis means 'logical explication'. The objective is not to provide a synonym of the *analysandum*, or even an expression with the same necessary and sufficient conditions of application. Nor is it to identify the true constituents and form which it possesses underneath the grammatical surface. It is rather to furnish an alternative expression or construction which serves the cognitive purposes of the original equally well, while avoiding its scientific or philosophical drawbacks. (Glock 2008: 159)

So one can distinguish between at least four forms of *analysis* that appeared in analytic philosophy after the linguistic turn: (1) *logical*, (2) *conceptual decompositional*, (3) *conceptual connective*, and taking the form of (4) *logical explication* or *construction*. The list, of course, is not exhaustive, but it should suffice to give rise to the question whether there is a clear-cut notion of "analysis" that covers all four forms. In Glock's view, there *may* be, yet a

vague one. In his opinion, in all of the above cases, "certain kinds of sentential paraphrase, formal or informal, still play a central role, and so do considerations about the applicability or nonapplicability of concepts to certain cases" (ibid.).

In fact, Strawson's colleague Paul Grice has described conceptual analysis by reference to the idea of "considerations about the (non-) applicability of concepts" (Grice 1989: 174), as does one of his followers Christian Nimtz. According to Nimtz, "what we are concerned with in conceptual analysis is *what our terms mean*" (Nimtz 2009: 138), and in his view, when one practices conceptual analysis, one engages in what he calls "scenario-based reasoning".

At this point, one could get back to the analysis of knowledge. This is Nimtz's description as to how the tripartite (JTB) analysis of knowledge gets rejected:

> Consider e.g. this line of thought:
>
> (1) Knowledge is not justified true belief. Just contemplate the scenarios Gettier (1963) puts forth. In the situations Gettier describes, we find a protagonist having a true justified belief that *p* – but he still does not know that *p*.
>
> This is a paradigmatic instance of what I call *scenario-based reasoning*.
>
> (Ibid.: 137)

Obviously, "we find" here does not mean *we find out empirically*, but rather *we consider it as a possibility*. Accordingly, the conclusion that the protagonist *does not know* is not a result of empirical discovery, but a result of our reasoning. As Glock notes, "during the heyday of the linguistic turn, puzzle cases were invoked to ascertain 'what we would say' under certain circumstances, in order to delineate the rules governing the use of philosophically contested terms" (Glock 2008: 164). So one of the central aims of conceptual analysis is to delineate the rules governing the use of our terms, or, as Nimtz prefers to put it – to make explicit the conditions that implicitly guide our application of the relevant term. Hence, in Nimtz's view,

in conceptually analysing an expression such as 'knowledge', we aim for an illuminating general characterization of the conditions under which it—or rather, the predicate '*x* knows that *y*'—applies across possible situations. So we contemplate possible cases until we come up with an analysis of the form "*A* knows that *p* iff *Φ*" where the right hand side makes explicit the conditions implicitly guiding our application of 'knows' all along. (Nimtz 2009: 138)

Nimtz here clearly follows the vein of Ryle. Ryle thought that philosophers should restate many of their traditional questions about mentality, and concede that the proper subject of investigation are the *concepts* which constitute our idea of mentality (concepts, for example, of believing, intending, understanding, thinking, wishing, being proud, generous, or vain). In the same way Nimtz seems to hold to the view that the proper subject matter of the analysis of knowledge should be described by making a "semantic-ascent": thus, the analysis of knowledge is a *conceptual* analysis of the *concept* of knowledge. Of course, in Nimtz's case it is not conceptual analysis in all of the senses that this term might have, but it surely falls under the vague concept of this notion.

Nimtz also adds an important proviso, which essentially derives from Wittgenstein's remarks on "family resemblance" concepts, and is relevant to what has been previously said about the possible routes that the analysis of knowledge has taken. He points out that "the conditions implicitly guiding the application of our terms typically aren't *Socratic*—i.e., they cannot well be captured by a tidy conjunction of individually necessary and jointly sufficient conditions" (ibid.).

It is true that in general we neither come to understand concepts by matching the words we use with necessary and sufficient conditions for their application, nor do we require that a competent user be able to provide such conditions. As Wittgenstein has pointed out, "When I give the description: 'The ground was quite covered with plants' – do you want to say I don't know what I am talking about until I can give a definition of a plant?" (Wittgenstein

1967: §71). And surely it is not just "plant" – one can easily add "knowledge", "intention", "meaning", "action", "causation" to the list, among others. So it is true that the conditions implicitly guiding the application of our terms typically are not Socratic.

However, this need not imply that when analyzing (the concept of) knowledge, intention, or meaning, one cannot aim at providing an analysis of X in terms of strict necessary and sufficient conditions. Rather, the lesson might be that one should increase the requirements for the critical evaluation of the proposed analysis. So stating a few counterexamples to the analysis might not be sufficient for its refutation, and one could require more of attempted refutations than just a few counterexamples. For example, a more general argument showing that the picture implicit in the analysis to be fundamentally mistaken.

On the other hand, a more general argument might be developed if one decides – at least temporarily – to forego an analytic definition, and focus more on the role that the relevant concept takes in actual practices, as was suggested by Oswald Hanfling (2000) and Edward Craig (1990) in the case of the analysis of knowledge. But again, that need not be taken to imply that *no* analytic definition is possible. For example, Wittgenstein, Ryle and White all paid close attention to the way that the concept of knowledge is used in our practices and ended up questioning the central idea underlying the tripartite analysis of knowledge – namely, that knowledge is an "an elite suburb of belief" (Ryle 1974: 5). Neither of them proposed, or aimed to propose, an analytic definition of knowledge. But the rejection of this central idea became an inspiration for John Hyman, who advocates the revisionary analysis of knowledge that *does* take the form of necessary and sufficient conditions. In his view, the correct theory of knowledge is: "A knows that *p* if and only if the fact that *p* can be A's reason for doing something" (Hyman 2006a: 908). In an explanation of what this statement amounts to, Hyman proposes the following:

In the generally accepted sense of the term 'analysis', the proposition

that A knows that *p* if and only if the fact that *p* can be A's reason for doing something is evidently an analysis of the concept *knows*, because it provides necessary and sufficient conditions for knowing that something is so without making use of the term 'know' itself or a cognate term. (Ibid.)

But he also adds two important provisos. First, he notes that it is not an analysis of the concept of *knowledge* (or *knows*) into more basic concepts, and so it is not possible to grasp the concepts expressed in the analysans without grasping the concept of analysandum. Second, he maintains that the theory of knowledge he advocates

does not pretend to provide a complex expression synonymous with the verb 'knows'. Nor does it imply that sentences of the form 'A knows that *p*' and 'The fact that *p* can be A's reason for doing something' have the same meaning. Indeed, it is obvious that they do not have the same meaning. For example, 'The fact that it is raining can be Pierre's reason for doing something' is not a paraphrase of the sentence 'Pierre knows that it is raining'; and a faithful translation of the second sentence into French will not include the word 'raison' or the phrase 'quelque chose'. (Ibid.)

Rather, Hyman writes,

what the theory does say is that 'A knows that *p*' and 'The fact that *p* can be A's reason for doing something' are logically equivalent; and that this equivalence is explained by the fact that knowing that something is so and being able to be guided by a certain fact are one and the same state. (Ibid.)

Hyman's case of the analysis of knowledge is a special one not only because it aims to defend a revisionist account of knowledge (or the concept of *knowledge*), but also because it seems to be a piece of *conceptual* analysis of a special kind.

So Hyman's analysis appears to have the following characteristics. First, (i) it is a piece of analysis of the concept *knows* because it takes a biconditional

form and provides conditions for *knowing* (vs conditions for the application of *knows*, at least seemingly so). However, (ii) it does *not* pretend to provide a synonymous expression or an expression that has the same meaning as that of analysandum (so analysans and analysandum are *not* co-intensional, or so it seems). Furthermore, (iii) it only implies that the two predicates are *logically equivalent* where the equivalence is explained by the fact that predicates apply to one and the same state. Given (i)-(iii), the question is: does Hyman's analysis state a (merely) conceptual truth or does it state a (genuinely) metaphysical one? Should the idea that the proposed analysis is not meant to result in synonymous expressions mean that it is *not* conceptual and that it specifies the *nature* of knowledge instead? Or should it rather be taken to mean that it *is* conceptual, that it *is* about the concept of knowledge, but that it does *not* require intensional equivalence of the analysandum and the analysans (or maybe that it *does* amount to intensional equivalence *without* having the same meaning)?

One can take a step back and put Hyman's analysis in the context of what has been said about the role that the empirical inquiry or discoveries might have for conceptual truths. It seems that Hyman's account of knowledge does not depend on empirical discoveries of the natural sciences, but is it likely that it is completely independent from *any* empirical facts? It does not seem to be a subject to verification or falsification by experiment or observation, and neither does it look like a truth that was discovered by observation. So the case is not an easy one, and that is only one example of what analysis might look like. But irrespective of the answers to the raised questions, the core question is this: is Hyman's analysis merely about how we conceive of knowledge (a conceptual truth, a piece of conceptual analysis), or is it about knowledge itself, or the nature of knowledge as it is in itself (a metaphysical truth, a piece of metaphysical analysis)?

But it is hard to believe in there being a true opposition, and here one might quote Peter Hacker's reply to a common objection to conceptual analysis in order to see why it is likely that the opposition is false. The objection runs as

follows: "conceptual analysis is about concepts, but we are interested in the nature of things, not in concepts". But as Hacker points out,

> the expression 'about' may look like a critical lever, but it is made of plasticine. A logico-grammatical investigation is no more *about* words than it is *about* concepts, and it is no more *about* concepts than it is *about* the nature of what is signified by the concepts. (Hacker 2013: 459)

Hyman also seems to be skeptical of there being a genuine opposition between an inquiry being about concepts *rather than* the phenomena we use concepts to think and talk about. He writes:

> Understanding the phenomena—science, in the broadest sense of the word—is a complex achievement, which depends on a number of different activities: devising theories, testing them experimentally, inventing and making scientific instruments, devising the mathematical and computational techniques which are used to develop theories and interpret experimental data, and inventing the new concepts and understanding the existing concepts in which theoretical ideas are expressed. (Hyman 2017: 309)

So in order to engage in what Hyman calls "science, in the broadest sense of the word", and thus aim to understand the phenomena, one must invent new concepts and understand the ones that are employed in expressing one's theoretical ideas. Concepts are *about* phenomena, and so an inquiry that is *about* concepts – a conceptual analysis, in the broadest sense of this notion – is at the same time an inquiry about the phenomena itself. Accordingly, a description of the *application-conditions* of the relevant concept or an explication of rules guiding the use of the relevant term, is *also* a description of what they apply to. So being about the concept of X does not imply that it is not about X or that it is not about the nature of X. Hence, conceptual analysis can perfectly well be seen as being an analysis of the nature of X, and there is no need for the explanatorily basic notion of "metaphysical necessity".

At this point it might be suggestive to add another perspective aiming to characterize the relation between the concept of X and the nature or essence of X. For sometimes, when a distinction is drawn – say, between event and process, or knowledge and belief, or mind and body – it is being asked whether the distinction is *merely* linguistic or semantic, or is it a *genuinely* ontological one. But it is not obvious that one must choose between the two, and the relation between semantics and ontology might be a closer one that it might seem. For example, David Wiggins, when speaking of the individuation of substances, writes:

> For someone to single out a leaf or a horse or a sun or a star, or whatever it is, that which he singles out must have the right principle of individuation for a leaf or a horse or a sun or a star … For to single out one of these things, he must single *it* out. Such truisms would scarcely be worth writing down if philosophy were not driven from side to side here of the almost unnegotiable strait that divides the realist myth of the *self-differentiating object* (the object which announces itself as the very object it is to any mind, however passive or of whatever orientation) from the *substratum* myth that is the recurrent temptation of bad conceptualism. It is easy enough to scoff at *substratum*. It is less easy to escape the insidious idea that there can be the singling out in a place of a merely determinable space-occupier awaiting incongruent or discordant substantial determinations (individually inconsistent answers to the question *what it is*). (Wiggins 2001: 150)

So Wiggins' point seems to be this: neither the anti-realist conceptualist, who maintains that the distinctions that *we make* are entirely the products of a conceptual scheme imposed on the world by *us*, nor the anti-conceptualist realist, who holds that the world exists already "chunked up" and nothing whatever remains for *us* to do except to perceive or discover what something is independently from any minds or concepts are right. Furthermore, both positions seem to be based on a misguided conception of the relation between

world and mind. Helen Steward – who shares Wiggins' position on this – notes that the truth of the matter might be that

> there is a *reciprocal* relation between world and mind: the things we single out—leaves, stars, and the rest—are indeed 'there already' and they have the properties they do quite independently of anything we might think or say about them; but, equally, we single out only *what we single out*—and what those things are is determined by the principles of individuation embodied in the sortal terms ('leaf', 'horse', etc.) by means of which we say what it is we are talking about in the first place. (Steward 2013: 800)

So an answer to the question whether any X-Y distinction is a distinction between two different ways of thinking or between two different kinds of entities might be not that it is either one of them, but that it is necessarily and unavoidably both. Consequently, an argument for an ontological distinction between X and Y can take the form of showing that there is a *conceptual* distinction between X and Y (between the concept of X and the concept of Y), and differential applicability of these concepts is consequently an argument for the ontological distinctness of the entities in question. Concepts and the nature or essence of things seem to go hand in hand, as do *conceptual* analysis and the analysis of the *nature* of X itself. Hence, constitutive claims *might* be said to be metaphysically necessary, but there is no good reason to believe that metaphysical necessity is an explanatorily primitive notion and some good reasons to hold the view that the modality involved is essentially conceptual.

2. 4. Causation, supervenience, and grounding

There are at least two reasons to distinguish a constitutive explanation, claim or relation from an explanation, claim or relation of a causal kind. The first reason is that causal explanations do not state conditions under which something *is* X, let alone that something is X *in all possible worlds*. Causal explanations state causal preconditions for X, whose nature is *independently*

identified, and explain *why* or *how* X came about. However, it does not imply that a constitutive theory of X is necessary for a causal explanation of X to be made. When commenting on the relation between *how* or *why* vs *what* questions in science, Burge notes that

> science is more interested in finding explanations of how and why things happen than in asking about nature… Often good scientific work can proceed without answering constitutive questions correctly. Still, obtaining clarity about key concepts, and delimiting boundaries of fundamental kinds indicated by such concepts, can strengthen and point scientific theory. (Burge 2010: xv)

The second reason why constitutive explanations are not to be confused with causal explanations is this: even if the verb "cause" can stand for a variety of things, and so can function in the provision of causal explanations of different kinds (see Steward 1997: 139-140), this does not cancel the idea that causal relations can hold only between logically independent entities, be it objects, properties, events, or items of some other categorial kind. As John Campbell notes,

> we generally assume that for there to be a causal relation between X and Y, X and Y must be logically independent; they must be "distinct existences," in Hume's phrase. If X and Y are not logically independent, then any statistical correlation we might find between them will reflect only that logical connection, and not a causal relation. (Campbell 2011: 234)

Given that constitutive explanations or truths explain the nature or essence of X, yet what is referred to in an explanandum is not what is logically independent from what is referred to in an explanans, it follows that constitutive explanation, truth or relation is not the same as an explanation, truth or relation of a causal kind. More formally, the argument runs as follows. If Y is what explains the nature of X (say, if the nature of knowledge is that it is an ability of a special kind, as it has been argued by Hyman), then X and Y are not logically independent. But if X and Y are not logically independent,

then Y cannot causally explain X (the ability does not causally explain knowledge; it *is* knowledge). Hence, if Y is the nature of X, then there is no causal relation between Y and X, and explanations or truths that concern the nature of X are not causal explanations.

When explaining the difference between the causal and the analytical (i.e. constitutive or conceptual) sense of "because", Hyman writes:

> "because" in "an apple is red because of the experience it produces in us when we see it" is treated as an analytical "because" by the philosophers who accept this proposition. That is to say, the proposition as a whole is about what kind of property redness is. For example, "an apple is red etc." is not meant to explain what causes an apple to be red. It is meant to express the idea that phenomenalism about colors is true: roughly speaking, that colors are the possibilities for experiences that bodies, surfaces, and light possess. (Hyman 2006b: 53-54)

If constitutive explanations or theories are not meant to indicate the causal preconditions of the phenomenon in question, then constitutive explanations do *not* compete with causal ones. Hence, Fodor is right: "*both* the causal *and* the conceptual story can be simultaneously true, distinct answers to the question of the form: 'What makes (an) *x* (an) *F*'" (Fodor 1975: 8).

But it is not only causal explanations or relations from which constitutive accounts or relations ought to be distinguished, although that might suffice in many cases. In contemporary analytic philosophy there are at least two other concepts that are being used in order to specify the relation between some X and some Y, and which therefore need clarification and comparison.

The first concept is the concept of supervenience. In recent philosophical literature, the term "supervenience" has been primarily used in formulations of non-reductive forms of physicalism. In general, if it is maintained that all non-physical facts about the world or its properties *supervene on* physical facts or properties, then the "is" in the subsequent conclusion that "everything is physical" should *not* to be interpreted as expressing a strict identity. Rather, it should be taken as signifying a form of *dependence*.

The notion of "supervenience" was first employed in its contemporary sense by R. M. Hare (1952: 145), who used it to characterize a relationship between moral and natural properties. But it was Donald Davidson, who introduced the term in the area of philosophy of mind. In his classic paper "Mental Events", where he defends a position known as "anomalous monism", Davidson writes:

> Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect. (Davidson 1980: 214)

On a general level, supervenience is commonly described as a relation that holds between two classes of properties. For example, if B ("base class") and S ("supervenient class") are two classes of properties with members b_1, b_2, …, b_n and s_1, s_2, … s_n, then S supervenes on B if the presence or absence of properties in S is "completely determined" by the presence or absence of properties in B. The notion of "complete determination" is supposed to mean that there cannot be any differences in S without differences in B, as well as that if all non-physical facts about our world supervene on physical facts, then "if you duplicate our world in all physical respects and stop right there, you duplicate it in all respects" (Jackson 1994: 160).

It is common to distinguish between the two forms that supervenience might take – weak vs strong. According to Jaegwon Kim, "A is said to *weakly supervene* on B just in case:

(I) Necessarily, for any x and y, if x and y share all properties then x and y share all properties in A – that is, indiscernibility in B entails indiscernibility" (Kim 1987: 315).

Whereas "A is said to *strongly supervene* on B just in case:

(III)   Necessarily, for any object x and any property F in A, if x

has F, then there exists a property G in B such that x has G,

and *necessarily* if any y has G, it has F" (ibid.: 316).

Three things shall be noted at this point. First, strong supervenience entails weak supervenience, but not vice versa. Typically, when talking of supervenience, philosophers appeal to its weaker form. Second, reduction[3] entails supervenience, but non-reducibility does *not* entail non-supervenience, because supervenience does not entail reduction. Hence, supervenience is a weaker relation than reduction. Third, if X supervenes on Y, then what follows is that Y is sufficient for X, but it does *not* follow that Y is necessary. But if Y is constitutive of X, then Y *is* necessary for X. Hence, constitutive relations and explanations should be distinguished from relations and explanations that are couched in terms of supervenience.

For some the idea of supervenience seemed to be too coarse-grained to deliver a satisfactory explication of the relevant kind of dependence or determination. Consequently, the idea of a seemingly different kind of dependence or determination has been introduced – namely, the idea of *grounding*. Irrespective of whether there is any difference between the two ideas, both supervenience and grounding are tied to, and partly motivated by the view that reality is structured in layers. It is held that objects, properties or facts of the "upper-level", *depend*, are *necessitated* or *completely determined* by the corresponding entities from the "lower-level" which is assumed to be more fundamental. As Louis deRosset puts it,

> reality comes in layers. We often disagree about what there is at the
> bottom, or even if there is a bottom. But we agree that higher up we find
> facts involving a diverse array of entities, including chemical,
> biological, geological, psychological, sociological, and economic
> entities; molecules, human beings, diamonds, mental states, cities, and

---

[3] There notion of "reduction" is used here in a way that implies modal equivalence: if a fact P is reducible to a fact Q, then it is necessary that P iff Q; if property F is reducible to property G, then it is necessary that *x* is F iff *x* is G.

interest rates all occupy higher layers. The nature and existence of the entities in the higher layers are determined by, dependent upon, and derived from the more fundamental facts and entities we find lower down. (deRosset 2013: 1)

Contrary to a rather technical term "supervenes", the notion of "ground" is widely used in ordinary (i.e. non-technical) language. In some cases, the notion captures the idea of causal dependence; in others, it is synonymous with the notion of "justification" or "reason". However, usually it is proclaimed that neither of these ideas are what is meant by the use of the notion of "ground" in the contemporary debates of metaphysics. But it is not that clear what *is* meant instead.

According to Gideon Rosen, "we should grant immediately that there is no prospect of a reductive definition of the grounding idiom: We do not know how to say in more basic terms *what it is* for one fact to obtain in virtue of another" (Rosen 2010: 113). But there being no prospects for a reductive definition, or no definition in "more basic terms" should not imply that no elucidation of the notion is possible. In general, the idea of grounding is being introduced by reference to examples by means of which one is expected to get a grip what is meant by the concept. Here are some of them: "the fact that the ball is red and round obtains *in virtue of* the fact that it is red and the fact that it is round" (Fine 2012: 37); "the dispositions of a thing are always *grounded in* its categorical features" (Rosen 2010: 110); "if an act is wrong, there must be some feature of the act that *makes* it wrong" (ibid.). There are more, although similar in kind. But the above examples should be sufficient to show that it is not clear whether there is any one kind of *determination* or *relation* that applies to them all.

Those skeptical of the idea of grounding (Hofweber 2009; Daly 2012; Wilson 2014) maintain that at least some of the proposed examples can be understood by using already familiar concepts: either that of conceptual relation, causal dependence or supervenience. But the believers tend to disagree. For example, deRosset maintains that "all proponents of grounding

agree that grounding relates facts, and that the facts that ground a fact are the facts that explain it" (deRosset 2013: 4). He also holds that "grounding explanations are not, or not just, causal explanations" (ibid.: 3, *f.*7) and "one entity may ground another, even though they are from disparate ontological categories" (ibid.: 4). So in his view, "a full specification of grounding relations among all entities would tell us how those entities "hang together" in something suitably like a layered structure" (ibid.: 2).

Meanwhile, according to Kit Fine, the notion stands for "a distinctive kind of metaphysical explanation, in which explanans and explanandum are connected, not through some sort of causal mechanism, but through some constitutive form of determination" (Fine 2012: 37). In his view, in the idea of constitutive form of determination

> it is not implied that the explanandum just *is* the explanans (indeed, in the case that there are a number of explanantia, it is clear that this requirement cannot be met). Nor need it be implied that the explanandum is unreal and must somehow give way to the explanantia. In certain cases, one might wish to draw these further conclusions. But all that is properly implied by the statement of (metaphysical) ground itself is that there is no stricter or fuller account of that in virtue of which the explanandum holds. If there is a gap between the grounds and what is grounded, then it is not an explanatory gap. (Fine 2012: 39)

But the idea that a constitutive form of determination implies that there cannot be a "fuller account" of that "in virtue of which the explanandum holds" is not that clear, and neither is the claim that there is no "explanatory gap" between the grounds and what is grounded.

Brian Epstein aims to explain the concept of grounding by means of the clarification of the notion of "determination". He writes: "When we say that G determines F, we sometimes mean a cause-and-effect relation: G causes F to be the case. Other times, we mean a grounding relation: G is a full metaphysical reason that F obtains" (Epstein 2015: 106).

Epstein is right that sometimes the notion of "determination", as well as the related terms like "depends on" or "because", *are* used to indicate a cause-and-effect relation, and it is sufficiently clear what that means in many cases. However, it is not obvious at all that in some cases where we use the notion of "determines" we mean a grounding relation that Epstein explains by reference to the notion of a "full metaphysical reason". And the reason why it is not obvious is it is not clear what the latter notion means. Epstein gives several examples that are supposed to express the same idea: "To say that G determines F is to say that G always makes F obtain. More specifically, G being the case guarantees that F is the case, and G is a complete metaphysical reason for F" (ibid.). So the idea of complete metaphysical reason is an idea of *always making something to obtain*, of *guarantee* (in some sense) of something being the case, or more generally, perhaps an idea of some kind of *necessitation* between things or facts that ground and things or facts that are grounded. It remains to be seen what kind of necessitation could possibly be indicated in such cases.

At least with respect to some examples, it seems that one *can* get some grasp on what might be meant by claims like "G always makes F to obtain" or "G fully determines F". Recall Rosen's claim that "the dispositions of a thing are always *grounded in* its categorical features" (Rosen 2010: 110). But what else can categorical features of dispositions be if not their causal-enabling preconditions: that is, conditions that *empirically* enable them to obtain. It is clear that dispositions and their categorical features "are from disparate ontological categories" (deRosset 2013: 4), and if they *are* from different ontological categories, then they are not the same. The crucial question then is what could it mean to say that dispositions are *always made to obtain* or are *guaranteed* by their categorical basis.

It seems that in general, if Y consists of a set of facts concerning all the relevant causal-enabling preconditions of fact X, then Y can "fully explain" or "determine" how it is empirically possible that X obtains. However, Fine maintains that if Y grounds X, then there is no "explanatory gap" between Y

and X (Fine 2012: 39). But that cannot be right if the concept of grounding concerns causal-preconditions of X, for a causal-enabling explanation is supposed to fill the explanatory gap: that is, explain why X obtains. On the other hand, if what is meant by there being no explanatory gap is that there is nothing to explain in why, given Y, X obtains, then Fine seems to be right in thinking that there is no good reason to believe in there being such a gap.

However, if what is meant by at least some uses of the idiom of "grounding" is X's dependence on its causal-enabling preconditions, then at least in some cases, grounding (as well as the supposedly different sense of "determination") *does* concern the cause-and-effect relation, even if "cause" or "effect" are not the notions that one would normally use in these contexts. For in cases when one inquires about the empirical preconditions of some facts, one means to ask about the causes of the phenomenon that is assumed to be their effect. Say, A wonders how it is possible (*empirically* possible) that B is able to see, and asks for an explanation of this fact E ("effect"). An explanation of E would presumably concern the neurological preconditions of this ability ("causes" or, more precisely, "causal preconditions"), and the facts C that concern these preconditions (ability's categorical features) would "fully explain" how the ability is possible. Perhaps one could say that neurological preconditions "fully determine" that the ability obtains, that there is no gap in this sense. But then, it should be obvious that if neurological properties constitute ability's categorical basis, then an explanation of the fact that E is grounded in C – a grounding explanation of E in terms of C – *does* concern the cause-and-effect relation.

Thus, it seems that the idea implicit in cases of "X is grounded in Y", when Y constitutes X's categorical features, is an idea of a cause-effect relation or causal dependence, where Y is a causal-enabling precondition of X, and Y fully explains/determines the empirical possibility of X. This shows that at least some "grounding explanations" are implicitly causal, and for this reason they should be distinguished from constitutive claims, as it has been argued at the beginning of this section. What the relation is between

constitutive explanations of X and grounding explanations that do *not* concern categorical features of X, and thus do not amount to causal dependence, ought to be decided on examining particular cases of the so-called grounding relation.

To conclude: constitutive relations, truths and explanations ought to be distinguished from relations, truths and explanations that are causal or based on the ideas of supervenience or grounding (when the latter concerns categorical features of the explanans). This does not imply that neither of them can contribute to a general theory of intentionality, for a general theory of intentionality includes different kinds of questions, and thus calls for different type of explanations. The basic lesson is simple, but important to bear in mind: an inquiry into the nature of X is *not* an inquiry into the causal preconditions of X, its supervenience base or its categorical basis, if there is any.

3. What is it to naturalize intentionality?

3. 1. Varieties of philosophical naturalism

"We are all naturalists now", Roy Wood Sellars declared in 1922. But as Neil Roughley pointed out years later, "the term can be used to designate anything from a broad commitment to keep the 'supernatural' out of philosophy to a methodologically highly specific conception of how that has to be done" (Roughley 2004: 48). Barry Stroud's semantic analogy seems to be apt in this case: he suggests to think of the notion of "naturalism" to be "rather like 'World Peace'. Almost everyone swears to it and is willing to march under its banner. But disputes can still break out about what it is appropriate or acceptable to do in the name of that slogan" (Stroud 2004: 22). However, indeterminacy notwithstanding, Jaegwon Kim rightly points out that "if current analytic philosophy can be said to have a philosophical ideology, it is, unquestionably, naturalism. Philosophical naturalism has guided and constrained analytic philosophy as its reigning creed for much of the twentieth century" (Kim 2003: 84).

Despite the fact that philosophical naturalism has been the predominant methodological approach in a large part of contemporary analytic philosophy, very little energy has been spent in explaining what naturalism is. As Mario De Caro and David Macarthur point out, "for the few who do take the trouble to explain naturalism, perhaps the most familiar definition is in terms of the rejection of supernatural entities such as gods, souls, and ghosts" (De Caro & Macarthur 2004: 2). This echoes Peter Hacker's view, who notes that "the most perspicuous use of the term "naturalism" in philosophy is when it is contrasted with super-naturalism" (Hacker 2011: 98). According to Hacker, in this sense of the term, both Descartes and Berkeley elaborated non-naturalist philosophies, and Kant's transcendental idealism was not a naturalist account of our knowledge: in the first case appeal to God was necessary to warrant knowledge claims concerning the natural world, and in the second, an appeal to *noumena* was deemed to be necessary. In the same sense Hume *was* a naturalist, because he explained the nature of knowledge without any recourse to any super-natural agencies. However, the contrast between naturalism and super-naturalism, does not play a central role in predominant contemporary forms of philosophical naturalism, and Hacker rightly notes that "those who marched under the banner of naturalism were not concerned with combating philosophical super-naturalism – they took its demise for granted" (ibid.).

Mark Timmons adds some positive content to what the modern version of philosophical naturalism amounts to. In his view, "the vague, pre-theoretic idea that the philosophical naturalist tries to articulate and defend is that everything – including any particulars, events, facts, properties, and so on – is part of the natural, physical world that science investigates" (Timmons 1999: 12). According to David Papineau, a necessary condition for naturalism is "the thesis that all natural phenomena are, in a sense to be made precise, physical" (Papineau 1993: 1), whereas David Armstrong maintains that "naturalism […] is the contention that the world, the totality of entities, is nothing more than the spacetime system" (Armstrong 1997: 6; similarly Katz 1990: 239) where "the

only particulars that the spacetime system contains are physical entities governed by nothing more than the laws of physics" (Armstrong 1997: 7).

So according to the common conception of naturalism, nature is to be conceived as a causally closed spatiotemporal structure that is governed by efficient causal laws, and human beings should be fully understood in terms of how they fit into this causal structure. Consequently, apart from being associated with the denial of the existence of God or the rejection of mind-body dualism, the notion of "naturalism" is commonly and primarily used "to mark one's acceptance of a scientific philosophy, or to denote the attempt to "naturalize" some allegedly contentious entities or concepts" (De Caro & Macarthur 2004: 3). Of course, that does not deny the fact that "naturalism means many different things to many different people" (Lawrence 2001: 1). But it does seem to suggest that, as De Caro and Macarthur rightly point out, despite its complexity and ambiguity, there is a substantial core conception of naturalism that underlies the majority of contemporary debates in analytic philosophy. In their view, the core can be represented as consisting of two main themes:

(1) *An Ontological Theme*: a commitment to an exclusively scientific conception of nature.

(2) *A Methodological Theme*: a reception of the traditional relation between philosophy and science according to which philosophical inquiry is conceived as continuous with science. (De Caro & Macarthur 2004: 3)

In order to distinguish this kind of naturalism from other, or older, versions De Caro and Macarthur suggest calling it "scientific naturalism". But the picture of its substantial core could be supplemented by two more elements.

First, the methodological theme is related to what is generally called "epistemological naturalism", according to which there is no genuine knowledge outside the scope of natural science. The view implies that it is *only* by following the methods of the natural sciences – or, at least, the methods of empirical a posteriori inquiry – that one can gain *genuine* knowledge. Second,

an ontological version of naturalism which stems from a commitment to a principle expressed in the "Ontological theme", is commonly accompanied with "a semantic version of naturalism that presupposes a prior commitment to the ontological version" (De Caro & Macarthur 2004: 7). For semantic naturalist, "the concepts employed by the natural sciences are the *only* genuine concepts we have and that other concepts can be retained if we can find an interpretation of them in terms of scientifically respectable concepts" (ibid.).

If naturalism is formulated as a thesis about what kinds of things exist, then it is *ontological* or *metaphysical* naturalism that is at issue. But one should take care: the "Ontological theme" not only says that natural science provides *a* true picture of nature; rather, it holds a more contentious claim, namely, that it is the *only* true conception of nature. As noted by many, ontological or metaphysical naturalism is based on Quine's idea that "it is within science itself and not in some prior philosophy, that reality is to be identified and described" (Quine 1981: 21), and is also expressed by Wilfred Sellar's idea that "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (Sellars 1963: 173).

It must be also emphasized that a commitment to scientific naturalism is not simply an expression of respect for the results of the natural sciences. Rather, a commitment to scientific naturalism involves a stronger claim: namely, that science is, or ought to be, the *only* genuine or unproblematic guide in matters of ontology, knowledge, method or semantics.

Scientific naturalism is the predominant form of philosophical naturalism, but not all contemporary philosophers who subscribe to naturalism accept the Sellarsian identification of what is real with what is describable by the natural sciences. Naturally, philosophers who belong to this group suggest distinguishing between scientific vs nonscientific forms of naturalism. For example, Peter Strawson suggests to distinguish between "soft", "catholic" or "liberal" naturalism vs "hard", "strict" or "reductive" one (Strawson 1986: 1-2, 38-41), John McDowell between "naturalism of second nature" vs "bald

naturalism" (McDowell 1996: 88-89), Jennifer Hornsby between "naïve naturalism" vs "orthodox kind of naturalism" (Hornsby 1997), and Barry Stroud suggests to distinguish scientific naturalism from what he calls a "more open minded and expansive naturalism" (Stroud 2004: 33). Although the ways in which these authors conceive of the alternative approaches differ in detail, all of them accept the idea that the reality is to be identified with what is natural (thus, there are no supernatural entities), yet reject the view that something can be natural and real only if it can be fully understood and explained in terms drawn from the natural sciences.

The dissenting voices are the minority, and it is the scientific kind of naturalism that is "*the* philosophical orthodoxy within Anglo-American analytic philosophy" (De Caro & Macarthur 2004: 8). Consequently, it is the orthodoxy that sets the agenda to what Terry Horgan and Mark Timmons call "naturalist programs". As they point out, the "naturalist programs"

> in philosophy are attempts to accommodate various kinds of discourse – e.g., moral discourse, mental discourse, mathematical discourse, semantical discourse – within a naturalistic worldview. These programs usually treat the vocabulary of natural science as relatively unproblematic, and various kinds of "higher – level" discourse as needing naturalistic accommodation. (Horgan & Timmons 1993: 182)

When speaking of the orthodox position – i.e. the scientific kind of naturalism – Glock points out that "all versions of naturalism come in both an eliminativist and a reductionist form" (Glock 2008: 139). So once faced with the seemingly non-natural phenomena like morality, modality or intentionality, a scientific naturalist "can either dismiss them as spurious or try to show that on closer scrutiny they boil down to a scientific or natural phenomenon" (ibid.). Consequently, if it turns out that the phenomenon cannot be fully explained and understood in terms drawn from the natural sciences, then the scientific naturalist concludes that – supernaturalism aside – the phenomenon does not exist. Thus, within the framework of scientific naturalism, either something can be given a constitutive account that would explain the relevant

phenomenon in terms of the fundamental categories of some natural science, or it must be dismissed as non-existent.

3. 2. Intentional antirealism and naturalization of semantics

For those who get involved in the "naturalist program" that is based on the agenda of scientific naturalism, the fork of eliminativism is not where one wants to end up, but what one wants to avoid at all costs. Of course, eliminativism might be a conclusion that one reaches *after* the failure of naturalization. But it is not what one can sensibly assume *before* the program.

In the context of the debates on the nature of intentionality, eliminativism takes the form of what gets called "intentional antirealism" or "irrealism". In the words of Stephen P. Stich and Stephen Laurence,

> intentional irrealism is the doctrine that meaning is a myth. A bit more precisely, it is the claim that nothing in the world instantiates intentional properties – that intentional predicates are true of nothing. If intentional irrealism is correct, then it is not the case that
>
> (1) 'Snow is white' means that snow is white.
>
> or that
>
> (2) George Bush often thinks about winning the next election.
>
> or that
>
> (3) Lincoln wanted to free the slaves. (Stich & Laurence 1994: 159)

More generally, eliminativism of the mental is taken to result in elimination of the so-called folk psychology, which would imply that psychological explanations are literally false. Of course, some philosophers have a taste to bite the bullet. But in this case, the majority's taste seems to be the opposite.

For example, Lynne Rudder Baker – one of the critics of scientific naturalism – holds that elimination of commonsense psychology is akin to a "cognitive suicide". In her view, "if a commonsense conception of the mental, in a wholesale way, may be mistaken, […] then the entire framework of terms of which we describe ourselves and others as desiring, hoping, fearing,

intending, supposing, speculating, imagining, hypothesizing, inferring, etc., may collapse" (Baker 1998: 1).

Meanwhile, Jerry Fodor – one of the prominent defenders of scientific naturalism – describes the general thesis of his book *Psychosemantics* as well as the costs of the alternative as follows:

> The main moral is supposed to be that we have, as things now stand, no decisive reason to doubt that very many commonsense belief/desire explanations are – literally – true. Which is just as well, because if commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species. (Fodor 1987: xii)

Hence, eliminativism better not be the main moral of a theory of intentionality or "psychosemantics", be it naturalistic or not. However, in case of scientific naturalism, the eliminativist position follows straightforwardly if intentionality or the mind cannot be "naturalized" in the relevant sense. But the sense needs explication.

Fodor uses the notion of "reduction" when explaining the idea of "naturalization" of intentionality:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of spin, charm, and charge will perhaps appear upon their list. But aboutness surely won't. […] It is hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. […] If aboutness is real, it must be something else. […] There is no place for intentional categories in a physicalistic view of the world. (Fodor 1987: 97)

The above passage is suggestive, but the notion of "reduction" seems to follow the same path as that of "naturalism" itself. As it has been rightly pointed out, the notion of "reduction" has taken a variety of forms, and thus "speaking of reductionism in general in contemporary analytical philosophy […] is not

viable" (Dagys et al. 2014: 152). Still, one can conclude that "naturalization" is a reductive agenda in some to-be-specified sense.

Robert Stalnaker maintains that "the challenge presented to the philosopher who wants to regard human beings and mental phenomena as part of the natural order is to explain intentional relations in naturalistic terms" (Stalnaker 1984: 6). However, as it has been shown before (see Part I, Chapter 2, section 2.4.), there are many different non-equivalent ways in which an explanation of some phenomenon might be given, so a requirement that an explanation should contain "naturalistic terms" does not suffice to decide which of them is intended.

Jason Stanley provides a more informative description: "a satisfactory naturalistic account of intentionality should explain *what it is* for a creature to have a belief or desire with particular content, without exploiting intentional notions" (italics – M.G.) (Stanley 2010: 89). An essentially similar description is also given by Fodor, who maintains that

> what is required […] is therefore, at a minimum, the framing of naturalistic conditions for representation. That is, *what we want at a minimum is something of the form "R represents S" is true iff C where the vocabulary in which condition C is couched contains neither intentional nor semantic expressions*. (Italics – M.G.) (Fodor 1984: 232)

Stanley also provides a description of what naturalization should amount to in case of informational semantics – one of the main theories of semantic naturalism. According to him, "the goal of an informational semantics is to provide a reduction of content bearing states to purely naturalistic notions by analyzing them as involving law-like, causal relations between agents and the environment" (Stanley 2010: 88).

Loewer gives a more detailed description of the desiderata that an informational semantics (IS) must meet. He writes:

> *(IS)* is intended as providing a *reduction* of facts about what A believes to physical facts, similar to, e.g., "Water is H_2O". In view of this, *(IS)*

must satisfy two adequacy conditions if it is to be a correct reduction. (1) *(IS)* and its instances must be law-like and true. The test of (IS)'s correctness is that the beliefs that it attributes must be tally with those attributed by folk psychology. […] (2) The states (Rn), the conditions C, and the notion of information employed must all be specifiable without appeal to semantic or intentional notions. (Loewer 1987: 288)

In case of biosemantics, Millikan is making it explicit that her aim is to provide a *theoretical* definition of mental representation (intentionality of thought, by another name). In her view, theoretical definitions should be distinguished from descriptive definitions, and the distinction between the two is explained in the following way:

Descriptive definitions are thought to describe marks that people actually attend to when applying terms. Conceptual analysts take themselves to be attempting descriptive definitions. Theoretical definitions do something else, exactly *what* is controversial, but the phenomenon itself, the existence of this kind of definition, is evident enough. A theoretical definition is the sort the scientist gives you in saying that water is HOH, that gold is the element with atomic number 79, or that consumption was in reality several varieties of respiratory disease, the chief being tuberculosis, which is an infection caused by the bacterium *Bacillus tuberculosis*. (Millikan 1995: 16)

Leaving aside the issue of whether Millikan is right about what conceptual analysis attempts, it is clear that even if Millikan does not see her agenda as aiming to "describe marks that people actually attend to", she is obviously interested in the same phenomenon as people, who attend to it by what she calls "descriptive definitions". In other words, the subject matter of Millikan's inquiry is not different: she aims to propose a definition of the phenomenon that gets picked out by concepts like *thinks that p*, *believes that q*, etc. Furthermore, she is also explicit in that she aims to provide a *definition* of representation. The central question of her book *White Queen Psychology and Other Essays for Alice* is "What is a mental representation?" (Ibid.: 1). Hence,

it is obvious that Millikan does not aim to provide a causal explanation of representation, or to specify its supervenience base or the base that "grounds" it.

Moreover, the distinctions between different kinds of definitions, and her interest in one of them, does not obscure the fact that her goal is to provide a constitutive theory of the nature of intentionality of thought ("mental representation"). This is clear from the examples Millikan gives as being cases of theoretical definitions (water, gold, consumption, etc.): they all are constitutive claims about the nature of the relevant phenomena. So the peculiar feature of Millikan's theoretical definitions is not that they aim to provide an essentially different account or explanation of the relevant phenomenon. Rather, their distinctive feature lies in them being restricted to a particular kind of vocabulary that can be used when aiming at what is essentially a constitutive theory of intentionality of thought or mental representation. The vocabulary is that of the natural sciences, and in biosemantics this means that it is the vocabulary of biology.

When discussing the possible grounds for defending constitutive claims, Jason Bridges maintains that

> informational semantics, and other forms of semantic naturalism, are probably best understood as defended […] on what we might call 'naturalistic grounds'. […] On the one hand, a naturalistic argument for a constitutive claim about f-ness will portray the claim as suggested by central features of our ordinary, pre-philosophical, pre-scientific understanding of f-ness. On the other hand, the argument will be shaped by a condition that will appear peripheral, at least at first glance, to our concept of an f – namely, the naturalist's constricted conception of the natural. (Bridges 2006: 529)

Consequently, the basic premise that guides the agenda of scientific forms of naturalization of semantics is this: "a constitutive claim about intentional content is that possession of content must consist in g, because g is the closest

anything comes in the 'natural' world to having the features and playing the roles that we pre-philosophically take content to have and to play" (ibid.).

However, sometimes the agenda of semantic naturalism is described in a somewhat different way. For example, several years later, Loewer describes it thus:

> Semantic Naturalism is a *metaphysical* doctrine about the status of semantic properties. […] If Semantic Naturalizers can find *naturalistic conditions that are metaphysically sufficient for semantic properties* and understand why they are sufficient they would show how semantic naturalism can be true and thus place the semantic within the natural order. (Italics – M.G., except for "metaphysical") (Loewer 1997: 109)

Meanwhile, according to Fred Adams,

> since the late 1970s and early 1980s there have been several attempts to naturalize semantics. While there are subtle differences between the various attempts, they share the view that minds are natural physical objects, and that the way they acquire content is also a natural (or physical) affair. […] *The goal is to naturalize meaning and explain how meaningful bits of nature arise out of non-meaningful bits. So we cannot rely on the meanings of words or intentions of agents to explain how thoughts acquire contents*. (Italics – M.G.) (Adams 2003: 143-144)

There might be two ways to understand the above descriptions of what semantic naturalism amounts to: either as implying a weakening of the traditional constraint on constitutive questions or as suggesting a change in questions of interest. However, both interpretations are problematic for the following reasons.

For one, there is no good reason for weakening the requirement on constitutive questions. A constitutive question of what it is for an expression to have a given meaning or for a thought to have a particular content ought to be distinguished from the questions of whether existence of semantic or intentional facts can be given a causal explanation by reference to non-semantic facts or whether it can be made consistent with physicalism. Causal

explanations of the empirical possibility of semantic or intentional facts presuppose that semantic or intentional facts exist, and do not even purport to answer the question of *what it is* for them to obtain. On the other hand, if all that one is interested in is a question of whether existence of semantic or intentional facts is consistent with physicalism, and the latter is formulated in terms of supervenience of semantic/intentional facts on the non-semantic/non-intentional facts, then, trivially, specification of the supervenience base (sufficient conditions) should suffice. But if one is interested in what constitutes having a thought with particular content, then there does not seem to be any good reason to hold that specification of sufficient conditions should be more important than specification of necessary conditions. Of course, it does not imply that sufficient conditions are irrelevant. The lesson is simply that a specification of constitutive conditions for X is not a specification of X's supervenience base.

Of course, the aim to specify the supervenience base or find out about the causal-enabling preconditions of semantic and intentional phenomena are perfectly plausible inquiries. However, the scientific form of semantic naturalism that is part of the agenda of naturalistic programs is generally taken to have a specific goal: namely, to show that the seemingly non-natural phenomena *are* natural (in the required sense of "natural") and *thus* avoid intentional antirealism. Nevertheless, neither of the above inquiries can show that intentional or semantic facts *are* natural. For if Y is a causal-enabling precondition of X, and Y is a natural fact or property (fact or property of some natural science), it does not follow that X *is* a natural fact or property, or that X is "reduced to" (i.e. is identical with) Y. The same holds for supervenience: if X supervenes on Y, and Y is a natural fact or property, it is perfectly possible that X is a non-natural (again, in the sense assumed by scientific naturalism) fact or property, and, by common view, that X is not reducible to Y. If that is sufficient for naturalization and avoidance of intentional antirealism, then this comes at a cost of the truth that not all facts and properties *themselves* are natural. In such a case, naturalization in this sense *could* avoid intentional

irrealism, yet it would imply a commitment to a kind of ontology that is non-natural by the standards of scientific naturalism. In other words, it comes at a cost of admitting that it is not the case that "if aboutness is real, it must be something else" (Fodor 1987: 97).

Be it as it may, the predominant versions of semantic naturalism are involved in what De Caro and Macarthur call a "substantive semantic project". As they point out,

> projects of naturalization have typically been conceived as substantive semantic projects in which the concepts of apparently nonnatural discourses must be: 1) reduced or reconstructed in terms of naturalistically respectable posits – that is, the posits of the natural sciences; or 2) treated as useful fictions; or 3) construed as playing a nonreferential or nonfactual linguistic role; or 4) eliminated altogether as illusory manifestations of "prescientific" thinking. (De Caro & Macarthur 2004: 4)

So as Bridges rightly points out, "informational semantics, as with any form of semantic naturalism, makes what we call a *constitutive claim*" (Bridges 2006: 529), and thus should be seen as aiming to provide a *constitutive* and *reductive* explanation of *what it is* for some mental state to have intentional content. It is to the analysis of the general theoretical framework guiding the predominant forms of semantic naturalism that the following chapter is dedicated.

II. Semantic naturalism: the standard view

1. Causal theories of mental content

Some version of causal theory of mental content has been the predominant approach to the problem of content determination in naturalistic theories of semantics[4]. But one should take note here: a causal theory of content determination is not a theory which says that causation or causal relations are necessary for having content or meaning. One might hold that without causation there would be no meaning or content, yet not conclude that it is causal relations that *determine* content. So thinking of causation as a necessary element for meaning or content is not sufficient for a causal theory of mental content. What the theory says is that causal relations *determine* content or meaning in the sense that A's V*ing* (belief, thought, etc.) *having* the content/meaning that it has (say, *p* as opposed to *q*) *consists in* these causal factors.

According to Fodor,

> only two sort of naturalistic theories of the representation relation that have ever been proposed. […] The two theories are as follows: that *C* specifies some sort of *resemblance* relation between *R* and *S*: and that *C* specifies some sort of *causal* relation between *R* and *S*. (Fodor 1984: 233)

Resemblance theories of the "representation relation" (i.e. of what determines *what R* represents, so what determines *R's* content) are widely considered to be inadequate for two reasons: first, resemblance, unlike intentionality or representation, is a symmetrical relation; second, one might "wonder whether resemblance is part of the natural order (or whether it's only, as it were, in the

---

[4] As Mark Greenberg notes, "it is confusing to talk of a "theory of content" since the theories in question are not theories of what contents are, but *theories of having content – of what makes it the case that thoughts have particular contents*" (italics – M.G.) (Greenberg 2005: 1, *f*.2).

eye of the beholder)" (ibid.). So it is generally thought that a theory of intentionality cannot be based on the idea of resemblance. Furthermore, the general line of thought is that a *naturalistic* theory of intentionality should be based on some version of causal theory of content. Fodor gives three reasons for this view:

> (1) Causal relations are natural relations if *anything* is. […] (2) Causation, unlike resemblance, is nonsymmetric, (3) Causation is par excellence, a relation among *particulars*. (Ibid.: 233-234)

However, as Robert Rupert notes,

> an appeal to causal relations does not, by itself, solve the problem of intentionality. Causes and effects permeate the universe, but intentionality does not. Thus, a CT [causal theories of mental content – M.G.] must identify the particular form or pattern of causal relations that establishes, determines, or constitutes an intentional relation. (Rupert 2008: 356)

It is a general feature of causal theories of content that they have trouble distinguishing between conditions for having content from the conditions under which those contents are true. As Fodor puts it, "the conditions that causal theories impose on representation are such that, when they're satisfied, *mis*representation cannot, by that very fact, occur" (Fodor 1984: 234). Hence, as Fodor points out, one of the main problems for causal theories of content is "that we want there to be conditions for the *truth* of a symbol over and above the conditions whose satisfaction determines what the symbol represents" (ibid.: 243).

A common solution to the problem of misrepresentation is based on the notion of "normal circumstances" (or its cognates). Here is Fodor's suggestion:

> Instead of thinking of the representation making conditions as whatever is necessary and sufficient for causing tokenings of the symbol, think of them as whatever is necessary and sufficient for causing such tokening *in normal circumstances*. (Ibid.)

67

In naturalistic theories of semantics, the notions of *normal circumstances* or *normal conditions* are meant to refer to all sorts of conditions that concern the agent's internal sub-personal system *and* the external natural environment. The notion of *normality*, in turn, is usually explained by means of the notion of *function*: roughly speaking, conditions are said to be normal if and only if a system can perform a function that it has been selected for. Appeal to normal conditions and functions is at the core of teleological theories of mental content.

It is common to characterize teleological theories as being of a different kind from causal theories. But the question whether this is true depends on whether causal theories of content can include teleological notions. One surely could stipulate that causal theories exclude teleology. However, it would run counter to the historical development of causal theories of meaning. For example, Dennis Stampe – one of the earliest advocates of a causal theory of meaning – holds that the meaning or content of some linguistic expression is determined by what would cause it if "fidelity conditions" (Stampe 1979: 88-89) were intact. Which of the conditions are fidelity conditions depends on the *function* of the item in question. Thus, theories of meaning or content determination that employ the concepts of *function* or *normal conditions* alongside the concept of *causation* can perfectly well be seen as causal theories of mental content.

This broader use of "causal theories of content" is in line with Robert D. Rupert's classification. As he points out, the general aim of theories as diverse as Jerry Fodor's "crude causal theory" (1987), Fred Dretske's informational semantics (1981) or Ruth Millikan's biosemantics (1984), among others, is to "identify the particular form or pattern of causal relations that establishes, determines, or constitutes an intentional relation" (Rupert 2008: 356). Thus understood, causal theories of mental content constitute a subclass of philosophical theories where the concept of *causation* plays a central explanatory role.

When giving a general overview of various causal theories that have been defended in analytic philosophy, Steven Davis writes:

> In the last 20 years or so, philosophers in the analytic tradition have taken an increasing interest in causal theories of a wide range of traditional philosophical topics. […] It is quite clear that causal theories of action, knowledge, memory, and perception contain references to mental phenomena. In each case the causal relation that is hypothesized to account for the relevant phenomena is a relation between some sort of mental object and some non-mental object. (Davis 1983: 1)

William Child proposes a more specific description of what the causal approach to the nature of mind amounts to. He writes:

> It is a commonplace of contemporary philosophy of mind that our ordinary, common-sense understanding of psychology is a form of causal understanding; and that many of our common-sense psychological concepts have an essentially causal element. So, for example, causality figures in the concepts of perception and memory: if S sees x, then x *causally explains S*'s perception; the same goes, *mutatis mutandis*, for memory. It figures in determining the contents of propositional attitudes, for, at least in certain central cases, the content of a belief is partially determined by the *normal cause* of beliefs of that type. (Child 1994: 90)

If one follows Child's characterization, then the central idea that follows from causal theories of mind is that causality figures in many mental *concepts*, such as perception, memory or thought. This idea is what is supposed to explain the deduction: if A V*s* that *p* or *x*, then it follows that *x* or the state of affairs signified by "that *p*" causally explain A's V*ing*. Thus, causality is part of the relevant mental concepts, or, alternatively, it is constitutive of the mental phenomena that these concepts are about.

Child's point that various causal theories of mental phenomena concern "our common-sense psychological concepts" should be kept in mind. For it implies that they aim to either clarify or reform (depending on one's point of

view) psychological concepts that are *ordinarily* employed in primarily (though not exclusively) non-scientific discourse. So these theories do *not* aim to define new concepts that are better suited for scientific or other type of inquiries. Of course, this does not rule out the possibility of modifying these concepts, if needed, or eliminating them if they turn out to be incoherent or useless. The important point to note here is that the theory is about *ordinary* psychological concepts[5] – as well as their subject matter, of course, but this is a false opposition – however vague, structurally complex, or incoherent they might turn out to be.

Causal theories of memory shall not be examined here, although brief or longer discussions of causal theories of knowledge, perception or action will be significant in the following sections. For now, the focus is on causal theories of content determination – more specifically, on the three problems that causal theories of this kind tend to face.

The first problem for causal theories of content is the problem of fine-grainedness of semantic properties. As Loewer points out, "semantic relations are apparently more fine-grained than causal relations" (Loewer 1997: 112). For example, S might think that *a* is *F* without thinking that *a* is *G*, and vice versa. But in cases where F *is* G, whatever causes S to V that *a* is *F* is the same as that which causes (or would cause) S to V that *a* is *G*. Given that causes are the same, yet one might think *a* is *F* without thinking *a* is *G*, it follows that whatever causes an agent to hold some intentional attitude is not sufficient to discriminate between the contents. Hence, causal relations are not sufficient for the determination of content.

The same problem is sometimes described by differentiating between two *types* of content: *informational content*, which is supposed to be determined by mere covariance relations, and *representational content* which is supposed to

---

[5] The term "ordinary use" is ambiguous, for as Gilbert Ryle noted (1971: 301-304), it may either refer to the *everyday* use of an expression (as opposed to a technical use of it), or the *standard* use of an expression (as opposed to a non-standard use of it). In this context, the notion of "ordinary psychological concept" should be understood as referring to a set of concepts that are expressed by mental terms (such as "belief", "desire", "intention", "thought", etc.) employed in a *standard everyday* use.

be semantic, and be the kind of content which is ascribed to propositional attitudes. However, as noted by Daniel D. Hutto, "even those who endorse psychosemantic theories in which informational content plays a foundational role accept that further conditions must be met before it can be regarded as representational" (Hutto 2008: 47). For example, Hutto writes,

> according to teleofunctionalists, informational content becomes representational if it is used to indicate or represent something for an organism. If and only if information-carrying states are assigned dedicated work in guiding organismic responses […] are semantic relations instantiated […]. (Ibid.)

However, the teleofunctionalists' solution does not seem to solve the fine-grainedness problem. For the idea that an informational content (which is too coarse-grained) becomes representational (which is *assumed* to be fine-grained enough) if and only if the former is used to indicate something only explains which factors turn informational into representational content. But mere reference to teleological factors (functions or the evolutionary history of these functions) does not explain *why* representational content *thus determined* is supposed to be more fine-grained than mere informational content. In other words, it is not clear why the use of causally determined informational content in guiding organismic responses should get one more fine-grained content than coarse-grained informational content. The mere idea that information-carrying states are assigned dedicated work in guiding action does not imply that the resulting content is more fine-grained than the kind of content which is determined by covariation relations and assigned to information-carrying states.

The second, more general kind of problem that any causal theory of content faces is that it fails to account for the phenomenon of semantic normativity. As indicated earlier, it is a general consequence of causal theories of content that they fail to distinguish between conditions that determine content-properties from conditions that would render that content correct.

The usual strategy to overcome the problem is to appeal to the notion of "normal conditions", and explain it in terms of notions like "proper-functioning", as is done in teleosemantic versions of causal theories of content. As Michael Luntley points out, "the central insight is that a state's responsiveness to some feature of the environment F will be selected just in case under normal conditions it will increase the survival chances of the creature that S responds to F" (Luntley 1999: 191). This strategy, however, is problematic for a theory that is based on a reductive kind of naturalism. As noted by Fodor, it is naturalistic only on the surface because the notions of "well-functioning" or "optimal conditions" are normative notions, even if *what they apply to* can be given an evolutionary explanation by reference to natural selection.

Recall that in its standard scientific form, semantic naturalism aims at a reductive explanation of intentional and semantic facts in non-intentional and non-semantic terms. Furthermore, a given explanation should not appeal to normativity, because one of the primary motivations for scientific naturalism is what Hilary Putnam calls a "horror of the normative" (Putnam 2004: 70). So normativity should either be eliminated or reduced. Consequently, teleosemantics, if it is to be a reductive theory, cannot appeal to well-functioning or optimal conditions without giving a reductive explanation of their nature. This is also noted by Stalnaker, who maintains that no "abstract account of information" which "takes normal conditions as given, can provide a naturalistic *reduction* of intentionality" (Stalnaker 1998b: 393). So if one aims to explain semantic normativity, then teleological explanation *might* be plausible, but that comes at the cost of turning into a theory of intentionality that is non-reductive, which for many implies that it is not naturalistic enough by the general standards of scientific naturalism.

The last problem relates to the phenomenon of self-knowledge, the main idea of which is well described by an example given by Brian Loar:

I am now attending to my thought that Freud lived in Vienna. I register

what the thought is about – Freud, Vienna, the one inhabiting the other. I note the thought's references and truth-conditions. I may be wrong about the non-semantic question whether Freud actually exists (timeless). But it is difficult to see how I might be wrong in my purely semantic judgement that this thought is about Freud if Freud exists. (Loar 1987: 96)

The phrasing "it is difficult to see…" should be interpreted in the context of causal theories of content which are generally taken to imply that what Loar calls a "purely semantic judgement" about an object of thought might be wrong. The idea that one might be wrong about the contents and objects of one's thought stems from the premise that the contents and objects of thought are determined by causal relations of which the subject is not aware: to use the phrasing of Tyler Burge, by factors that are "not part of the cognitive world of the believer" (Burge 1977: 359), or as Colin McGinn puts it, by something "external to what cognitively transpires in the mind of the thinker" (McGinn 1983: 68).

So it looks like causal theories of content have the following implication: if *q* is what causes A's V*ing* that *p* (even in normal conditions), then the content of A's V*ing* is in fact *q*, even if A says and thinks that it is *p*. Of course, self-knowledge does not imply that A cannot be wrong about the truth value of what A V*s*. Rather, it is a point about one's own ability to know *what* one V*s*, not to know the truth value of it. And it does seem that a causal theory of content is committed to denying that one has an authority (sometimes called "first-person authority") as to what one V*s*.

Contemporary versions of causal theories of content are sometimes presented as being inspired by Kripke's causal theory (or picture) of reference determination (Kripke 1980). However, as it is noted by Stalnaker, "Kripke, in his defense of a causal theory of reference, explicitly disclaims any reductive ambitions" (Stalnaker 1999: 206), so in this respect Kripke's theory is different from the ambitions of the reductive versions of causal theories of content. The main reason why Kripke's theory is non-reductive is that it makes an essential

use of the notion of *intention* in its account of reference determination: reference, in Kripke's view, is determined not by any epistemic facts (such as beliefs or knowledge), but by a combination of a specific kind of semantic intention *and* causal factors that obtain in the situations in which the speaker utters the relevant expression. On the other hand, there is also an important similarity between Kripke's causal theory of reference and reductive causal theories of content determination. According to *both*, reference or content is determined by *non-cognitive* factors (assuming that in Kripke's case, the notion of semantic intention is not constrained by epistemic considerations). In Kripke's case, it amounts to the idea that it is not a Fregean *Sinn* – by general agreement, an epistemic factor – that determines reference (a Fregean *Bedeutung*), but rather by a non-cognitive kind of semantic intention (if that makes sense) *and* causal relations. In reductive forms of causal theories of content, it translates to the idea that causal relations – by general agreement, not an epistemic factor – are both necessary and sufficient for the determination of content. In non-reductive forms of causal theories of content, it might translate into the view that certain epistemic factors which are part of the "normal conditions" are also necessary, which would imply that causal theories of this kind are different from Kripke's causal theory of reference in this respect.

There might be various reasons for adopting some version of a causal theory of mental content. One of the main reasons as to why a *naturalistic* theory of intentionality might find it necessary to appeal to causation has been already identified: "causal relations are natural relations if *anything* is" (Fodor 1984: 233-234).

But one might worry about the implication of this view. For if intentionality consists in some causal relation, then it *is* a relation, and this is not the conclusion that some would like to reach. For example, Brentano reasoned as follows:

If someone thinks of something, the one who is thinking must certainly

exist, but the object of his thinking need not exist at all. […] The terminus of the so-called relation does not need to exist in reality at all. For this reason, one could doubt whether we really are dealing with something relational here, and not, rather, with something somewhat similar to something relational in a certain respect, which might, therefore, be called 'quasi-relational'. (Brentano 1995: 272)

However, the idea that intentionality is a relational property, although of a peculiar kind, is what many philosophers assume rather than deduce some version of a causal theory of mental content. For example, Michael Tye writes that "the property of representing what does not exist is, on the face of it, a very strange property quite unlike other relational properties" (Tye 1994: 123). Robert Stalnaker too seems to be puzzled by the relational character of intentionality:

For various familiar reasons, intentional or representational relations seem unlike relations holding between things and events in the real world. […] One can, it seems, picture, describe, or think about such things as gods and golden mountains even if they do not exist. (Stalnaker 1984: 6)

So intentionality or content possession seems to be quite unlike other relational properties, but it *is* relational nonetheless. But why not conclude, with Brentano, that it is not a genuine relation or that it is a quasi-relation at best? For one, it seems that the conclusion that it is a quasi-relation is not the one that a naturalist would like to reach, for the class of quasi-relations that are part of the natural world is empty. But the alternative option – namely, the idea that it is *not* a relation – seems to be rejected not so much on the basis of the intuition that it *must* be a relation because it appears to be such, but from a set of specific theoretical commitments that are implicit in one's theorizing about the nature of intentionality and mental representation, and that manifest a picture of the mind which underlies the relational model of intentionality of thought.

2. The bifurcationist picture of the mind

Recall that according to causal theories of mental content, causal relations are what *determine* or *constitute* "intentional relations": it is what makes it the case, in the relevant sense of this term, that A V*s* that *p* as opposed to *q* or some other content. Furthermore, a causal relation that "is hypothesized to account for the relevant phenomena is a relation between some sort of mental object and some non-mental object" (Davis 1983: 1). The relevant "mental object" might vary, perhaps to the extent that it need not be "purely mental", but the idea is that a causal relation is supposed to account for the relation that obtains between one kind of object and other kind of object (never mind that one is mental and the other not).

If causal relation is being postulated to account for the relation between the two relata, then the relation between the two is being assumed. For otherwise there would be no need for it to be accounted for; that is, no need to specify what kind of relation obtains between the two relata. More specifically, if a causal relation is postulated to account for intentionality of thought (or possession of content), then it is assumed that intentionality or possession of content *is* a relation that needs to be accounted for. It might seem that the assumption is an innocuous one and has no philosophical importance. However, the assumption seems to manifest a picture of the mind (and its relation to the world) that is essentially Cartesian. The picture has been quite aptly characterized by John McDowell, who calls it by various names: "bifurcationist", "constitutively duplex", "hybrid", "the two component", or "the highest common factor" approach. But irrespective of differences in names, they all express the idea that McDowell describes in the following way:

> The idea is that part of the complete truth about the mind is the truth about something *wholly in the head*; another part of the complete truth about the mind is the truth about how *the subject matter of the first part is related to things outside the head*. (Italics – M.G.) (McDowell 1998a: 278)

Richard Holton's characterization might also be suggestive. As he sees it, "this approach involves treating such phenomena as perception, knowledge, memory, and the content of thought as composite: as consisting of different factors that can obtain independently" (Holton 2000: 1).

For present discussion, one should bear in mind that the notion of *the content of thought as composite* should not be taken to necessarily mean that there are two types of *content*, each of which can obtain independently from each other. It surely can be taken to mean this, although what "content" means might vary in different cases. For example, it might be the case that one *part* of the composite whole contains "content" that is not of the same kind as the one constituting the other composite part: say, a picture "in the mind" as well as what it is about might be called "content", so one should take care. But it might also be taken to mean that the notion of "having a content", or the notion of "V*ing* that *p*" more generally, is a composite notion.

It should also be added here that the bifurcationist picture of X does not result from some factor being necessary for X. If, say, Y is necessary for X, it does not follow that X is a composite notion. For example, causal relations might be necessary for perception or knowledge, but this does not imply that "seeing X" or "knows that *p*" are composite notions. What does seem to be true is that at least in the domain of the philosophy of mind, some notion of X is a composite one – or an account of X is a bifurcationist one – if X (or the concept of X) is analyzed in terms of two different and independent factors which are both individually necessary and jointly sufficient for X, where one of the two is "internal" or "wholly in the head" in some plausible sense of these terms, whereas the other is "external" and is supposed to relate the first one to the things outside the head (the external world).

Take, for example, the analysis of knowledge. The idea that an external relation to the environment is a necessary condition for knowledge does not imply or presuppose the bifurcationist account of knowledge. For it does not imply that the external relation is part of the two factors that are constitutive for knowledge. But once the analysis of knowledge takes the form of

something like a causal theory – a theory that analyzes knowledge in terms of some presumably internal epistemic factor (justified belief that *p*) *and* some external factor (say, a reliable causal relation to the external world), which are individually necessary and jointly sufficient for knowing that *p* – then the bifurcationist picture of knowledge is manifest.

Or take a case of perception – again, a *causal* theory of perception. John Hyman maintains that, according to the modern version of it, as it was developed by Peter Strawson and Paul Grice (Grice 1961; Strawson 1974, 1979), "the causal theory of perception says that 'X saw M' describes a psychological episode in terms of its cause" (Hyman 1993: 210). Strawson and Grice were explicit in saying that their theory is an analysis of the ordinary notion of seeing (Grice 1961: 143; Strawson 1974: 72). As Grice put it, the modern causal theory of perception aims "to elucidate or characterize the ordinary notion of perceiving a material object" (Grice 1961: 1-2). Accordingly, it is a constitutive account of perception, which mentions factors that are both individually necessary and jointly sufficient for perception (or perceptual experience). The theory says that perception should be analyzed in terms of some psychological episode and the causal relation that it bears to the environment. Furthermore, the psychological episode is supposed to be of a kind, "which it is possible to have even though nothing which corresponds to it is within sight (or within earshot)" (Hyman 1993: 210). So an episode is one of the two factors – the one which is "wholly in the head" in the sense that it might occur even when nothing corresponds to it – and the causal relation is another one which is supposed to relate the psychological episode to the external world when it *is* a case of perception. It should be clear that this is a kind of duplex or two-factor approach to the nature of perception.

It is perhaps an accurate description of any duplex conception of some mental phenomenon – be it thought, knowledge, perception – that it is a Cartesian one. For example, when commenting on the causal theory of perception, Hyman notes that

the psychology implicit in the causal theory is thoroughly Cartesian. As we have seen, Strawson explicitly rejects the proposition that the immediate objects of perceptual awareness are psychological entities of some sort. But the causal theory is still committed to the Cartesian illusion that 'the ordinary notion of perceiving' is a composite notion, which can be divided into its purely mental and its physical components, each of which can exist without the other. (Hyman 1992: 294)

More specifically, Hyman writes, according to the modern version of the causal theory of perception, "any perceptual statement, such as 'He has spotted a misprint on the first page', 'I heard the professor's inaugural lecture', etc., is equivalent to a report of a psychological episode plus a causal hypothesis which is meant to explain why this episode occurred" (ibid.: 293).

So according to Hyman, "the Cartesian illusion", which the causal theory of perception is committed to, consists in the idea, implicit in the theory, that the ordinary notion of perceiving is a composite notion, and that it involves a "purely mental" (i.e. psychological episode) and a physical (i.e. causal relation) component. This might be taken to mean that involving a "purely mental" component is necessary for the illusion to be manifest. But it is not obvious that the component must be "purely mental" for the two-factor approach to be present. For example, McDowell seems to agree that for "the inner realm" to be "purely mental" might be a necessary condition for the "classically Cartesian picture" (McDowell 1998a: 244), but he does not seem to agree with the idea that the "purely mental" component is necessary for the picture of the mind to be a bifurcationist one (or be a "non-classically" Cartesian one). According to him, in the contemporary materialist versions of the Cartesian picture,

the interiority of the inner realm is literally spatial; the autonomous explanatory states are in ultimate fact states of the nervous system, although in order to protect the claim that the explanations they figure in

are psychological, they are envisaged as conceptualized by theories of mind in something like functionalist terms. (Ibid.: 244)

Moreover, McDowell adds,

in the physicalistic modern version of the insistence on autonomy the self-standingness of the inner realm suffices to exclude intrinsic involvement with the world […]. The intrinsic nature of inner states and events, on this view, is a matter of their position in an internal network of causal potentialities, in principle within the reach of an explanatory theory that would not need to advert to relations between the individual and the external world. (Ibid.: 250)

For the above quotations to make sense, one must give up on the premise that the notions of "self-standingness" or being "wholly in the head" must imply some kind of mentalism or immaterialism. Be it as it may, a general idea seems to be that a bifurcationist picture of the mind, Cartesian at least in the non-classical sense, is a picture which portrays the nature of the phenomena in question as consisting of two factors that are independent from each other, which in the philosophy of mind translates to the idea of one factor being "wholly in the head" in some sense of this term, and the other being of a kind that relates the first one to the things outside the head.

The reasons why one ends up being committed to duplex accounts of mental phenomena might vary, and one should consider specific cases. But what does seem to be true more generally is that a bifurcationist account of some mental phenomenon follows *either* if one is committed to some preconception as to what the concept of the relevant mental phenomenon applies to (say, the concept of seeing an X or thinking that *p*) which then requires an account that would explain how the supposedly internal subject matter of the relevant concept relates to the external world, *or* if one is committed to some theory as to how some supposedly internal element relates to the external world which, in turn, rests on the assumption that the concept of the relevant mental phenomenon applies to something that *needs* to be related to the things outside the head. One should bear in mind that this does not put

too many constraints on the nature of *that*, which the concept of X picks out or describes, except that it must be "wholly in the head" and so of a kind that something else, most likely external to it, must be considered to be necessary for it to be related to the things outside the head.

The following sections are dedicated to a discussion and conceptual analysis of the theoretical framework that is generally presupposed by standard versions of semantic naturalism. The main aim of the analysis is to show that the general conceptual framework which is predominant in the contemporary philosophy of mind and which underlies the standard versions of semantic naturalism amounts to the bifurcationist or constitutively duplex picture of intentional states. The diagnosis is general in character yet restricted in its scope. For one, it does not aim to demonstrate that *all* versions of semantic naturalism *must* rest on the presuppositions that are to be explicated (a view that is hard to justify). Moreover, the proposed characterization does not aim to apply to those versions of naturalistic theories of semantics – if there are any – which do *not* assume the philosophical views that are to be explicated and used as a basis to show them resulting in a bifurcationist picture of the mind. On the other hand, there appears to be good reason to believe that the *standard* forms of semantic naturalism, such as Fodor's causal theory, Dretske's informational semantics or Millikan's biosemantics *do* rest on these assumptions, most of the time to the extent that they are not even mentioned explicitly, and are detectable only via the vocabulary that is used. For example, the common use of concepts like "internal physical state" playing some causal-functional role seems to provide a good reason to assume that a theory is committed to a functionalist account of mind. Accordingly, if a question about the nature of intentionality is phrased in terms of "how can a mental representation be about something or have content?", then it seems to be reasonable to assume that it is the representational theory of mind that guides one's theorizing.

In any case, the proposition that semantic naturalism is an implicitly Cartesian and bifurcationist theory is not an evaluation of *all* possible modes of semantic naturalism, but of a set of specific theoretical commitments that

appear to be implicit and guiding the theorizing of the predominant versions of semantic naturalism. More specifically, the diagnosis applies only to those versions of semantic naturalism which do – though need not – presuppose the following theories: (i) the standard conception of semantic externalism, and *either* (ii) the orthodox account of functionalism that is based on token identity theory and causal analysis of mental concepts, *or* (iii) the representational theory of mind and concepts. Although (ii) and (iii) do often go hand in hand, one of them seems to be sufficient – when combined with (i) – for a constitutive theory of intentionality of thought to be a Cartesian and a bifurcationist one.

3. Semantic externalism

3. 1. The externalist character of psychosemantics

It is rarely made explicit, but naturalistic theories of semantics are, as a general rule, committed to semantic externalism. The prominent semantic naturalist Fred Dretske writes as follows:

> Materialists should be willing to tolerate some degree of externalism about the mind. […] I know of no plausible psychosemantics, no plausible theory of what makes one thing *about* another, that isn't externalist in character. It is the relations – causal, informational, historical, or whatever – that, on a given Sunday afternoon, makes something in my brain about football rather than philosophy. Thoughts are in the head, but what makes them the thoughts they are is not there. (Dretske 1996: 143)

Two conceptual points to note before commenting on this passage. First, Dretske qualifies the notion of *materialists* in a footnote by adding that it concerns only those "who are realists (i.e., not eliminativists) about the mind" (ibid.: *f.*1). This, in turn, confirms the point, argued before (see Part I, Chapter 3, section 3.2.), that a naturalistic theory of psychosemantics – externalist or

not – is of no concern for eliminativists about the mind. Second, the idiom of "something in my brain" is something to bear in mind for the following section, as well as the idea that thoughts are in the head. In this quote, they are not very informative, yet suggestive. And in order to get clearer as to what they mean, one should interpret them against the background of a more general framework about the nature of the mind that remains implicit in the above passage.

The quoted fragment does not say this, but Dretske is a defender of informational semantics. However, from what is being written, it looks like he is more convinced of the idea that *some* kind of external relation must be a part of psychosemantics (causal, historical, or whatever), than that it must be informational in some, plausibly externalist, sense. So according to Dretske, even if informational semantics is not the true story about intentionality, then perhaps biosemantics is – a naturalistic psychosemantic theory that alludes to facts concerning the evolutionary history of certain brain functions. Or perhaps some other version of causal theory of content must be true, if none of the above are on the right track ("other", because informational semantics *is* a form of causal theory, as is teleo- or biosemantics, according to some classifications). One way or the other, naturalistic psychosemantics, according to Dretske, must be externalist in character. This is in fact a general view of the standard versions of naturalistic theories of semantics, and the commitment to externalism is implicit not only in Dretske's theory, but also in theories developed by Fodor (1994: 7; 1998: 14) and Millikan (1984: 268).

As noted by Fred Adams,

> at least since the mid 1970s, externalistic theories of content have urged that thought contents depend crucially upon one's environment, and do not depend solely upon what is inside the head (for most thoughts). […] What the naturalizers of meaning add to the picture of meaning externalism is a mechanism. (Adams 2003: 143)

More specifically, in Adams' view, the semantic naturalist adds

an account of the mechanism that explains how external physical objects become correlated with the internal physical states of one's head (mind) such that the internal physical states come to mean or be about the external physical objects. (Ibid.)

It is a substantial assumption whether what Adams calls "internal physical states" are such that they *mean* or *are about* external physical objects. The phrasing is suggestive in a way that it contributes to the general picture as to what "naturalizers of meaning" are aiming at. But for now it is not necessary to examine the assumption in detail. What *does* require a reminder though is that semantic naturalists are *not* in the business of furnishing causal explanations as to why some internal physical states are in a certain condition (whatever that might be), or why they correlate, if they do, to certain external stimuli. That is not a matter of philosophical speculation or theory, but a question for specific empirical sciences.

Accordingly, appeal to the what Adams calls "mechanisms" should not be taken to imply that psychosemantics is proposing an alternative empirical theory concerning the indicated questions: i.e. why certain brain states are in a certain condition, what happens in the brain when certain external stimuli are present, etc. For it seems to be plausible to assume that their actual aim, even if phrased in physicalistic vocabulary, is to provide a naturalistic answer to the content-determining question: namely, to supplement the general picture of semantic externalism – whatever it is – with a mechanism which would be naturalistic enough and could work as an answer to the question as to what determines the content of mental states (which are assumed to be indicated by the idiom of "internal physical states of one's head").

The nature of the indicated mechanisms is of no concern for present purposes, for the idea of bifurcationism, as explained before – more specifically, the idea of the "external" element that is constitutive for the relevant mental phenomena – does not seem to engender any specific requirements about the nature of these mechanisms: causal, historical, informational or whatever should do, as long as they are "external" in the

relevant sense. It is of great importance, however, to get clearer as to what is meant by a saying that a theory is "externalist in character", or what a commitment to the conceptual framework of so-called externalistic theories of content amounts to.

It is generally agreed that externalism "has many far-reaching philosophical consequences for the epistemology of mental states, first-person authority (about the content of a person's own mind), personal identity, skepticism, the nature of mental properties, physicalism, mental causation, and the role of content in scientific psychology" (Jacob 1991: 723). However, there is no *one* externalism, for the notion has been explained in several non-equivalent ways.

For example, Akeel Bilgrami characterizes externalism in a general way, and links it to the idea of a skeptical scenario turning out to be true. In his view,

> a general characterization of the doctrine of externalism is that the contents of an agent's beliefs are not independent of the world external to the agent. It is a denial of the view that intentionality is fully characterizable independent of the external world, or to put it in terms of Descartes's First Meditation, it denies the view that an agent's intentional contents would be just what they are even if it turned out that there was no external world. (Bilgrami 1992: 2)

Thus, according to Bilgrami's general conception of externalism, intentionality cannot be fully characterizable without making reference to the external world, and it also implies that it is dependent (in some sense) on the world external to the agent. It is not clear whether it implies that intentionality is impossible if there is no external world, but it does say that if it turned out that there were no external world, then it would not be the case that our "intentional contents would be just what they are" (whatever that means).

One way or the other, Bilgrami's conception of externalism does not help one to understand why psychosemantics must be externalist in character, or even give a sense as to what it might mean. The reason why it cannot explain

this is because it is not specific enough and reference to the skeptical scenario is of no conceptual use for this explanatory purpose.

Sanford C. Goldberg gives a more specific description as to the kind of views that the concepts of *internalism* and *externalism* imply. In his view, the internalism-externalism debate

> in the philosophy of mind and language concerns the *supervenience base of properties regarding linguistic meaning, mental content, and the propositional attitudes*. Hence the traditional formulation of psychological internalism is as the thesis that psychological properties supervene on the intrinsic, non-relational features of the physical body of the subject instantiating those properties; and the traditional formulation of psychological externalism, as the denial of this supervenience thesis. (Goldberg 2007: 2)

A similar description is offered by Sarah Sawyer, who characterizes the positions as follows:

> Internalism, then, is the view that psychological properties supervene locally on physical properties: no two individuals could differ psychologically without differing in some intrinsic physical respect. Externalism rejects this local supervenience thesis, maintaining in contrast that individuals could be exactly alike with respect to their intrinsic physical properties and yet differ psychologically – if, for instance, they were related to relevantly different environments. (Sawyer 2011: 133)

Both Goldberg and Sawyer characterize the distinction in terms of the supervenience base of psychological properties regarding linguistic meaning, mental content, or propositional attitudes more generally. More specifically though, internalism is defined in terms of *local* supervenience of the mental on the intrinsic, non-relational features of the physical body or intrinsic physical properties. Externalism is then defined in terms of the denial of internalism *thus understood*. This, in turn, amounts to the following positions:

SupInt. Psychological properties (e.g. believing that p, meaning that q, etc.) supervene locally on intrinsic/non-relational physical/mental properties, i.e. having "intrinsic" physical/mental non-relational properties is *sufficient* for having psychological properties.

SupExt. Psychological properties do *not* locally supervene on intrinsic/non-relational physical/mental properties, i.e. having "intrinsic" physical/mental non-relational properties is *insufficient* for having psychological properties, and relation to the relevantly different environment can make a difference as to what psychological properties the subject has.

This is more specific than a general conception of externalism, though it still lacks one important idea that is characteristic of semantic externalism, as it is generally understood and presented. However, reference to *intrinsic* properties being insufficient for having of psychological properties does seem to contribute to an understanding as to why one might appeal to external relations: if intrinsic factors are insufficient, then it seems to be natural to conclude that something "external" must be added.

However, **SupExt** is not specific enough about what these additional factors might be. It is true that **SupInt** implies that propositional attitudes (etc.) are determined by intrinsic physical/mental properties, and because of this **SupInt** captures what has been generally called "semantic" or "content internalism": a theory which provides an answer to the content determination question. But it should be obvious that **SupExt**, when taken on its own, does *not* provide an alternative answer to the same question. For **SupExt** only implies that linguistic meaning, mental content or propositional attitudes are *not* determined by what is "in the head", in the sense that what is "in the head" is insufficient to determine propositional attitudes (etc.). The reason why **SupExt** has no answer to the relevant question is because it is defined in terms of the negation of **SupInt**, which, naturally, does *not*, by itself, result in a positive account as to what determines propositional attitudes or their content.

One could think, however, that the mere fact that nothing "in the head" is sufficient for determination of content implies that the content is determined by something that is outside the head. But that does not follow. For there is no good reason to assume that it must be determined by either what is "in the head", or what is "external" to it at this point. What *SupExt* does allow to assume perhaps is that the relation to the external environment is necessary for having propositional attitudes (etc.). However, that still falls short of the relevant answer. For example, one might hold that it is necessary for A to be in the environment in which the game of chess is played in order for A to be able to *intend* to play chess (Wittgenstein 1967: §197, §337). So it might be true that the relation to the environment is necessary for A to be able to intend to play chess, and that the environment can make a difference as to what intentions one can have, although nothing is being said at this point as to how that difference is attained. But it should be obvious that if externalism is formulated in this way, it still says nothing as to what determines the content of A's intention or thought.

That is why both positions have been formulated in terms of specific views as to what determines content, and in most cases it uses neither the notion of supervenience nor the Cartesian possibility of skeptical scenario. For example, Glock and Preston provide what could be considered a standard characterization of the distinction. They write:

> According to *internalism*, the content of A's intentional attitudes (propositional attitudes, intentions) is determined by non-relational, internal properties of A's mind or brain. Although this has been the received view since Descartes, *externalists* have recently denied it. They maintain that what A thinks is at least partly determined by facts "external" to, and perhaps even unknown to, A, notably facts about A's physical or social environment. (Glock & Preston 1995: 515)

Reference to "external" facts, which need not be (and generally are not) known, is accurate, although the specifics as to what those facts are might differ in different versions of semantic externalism. For example, Donald

Davidson has argued that historical properties are essential for meaning (Davidson 1987a: 447), Tyler Burge has argued that it is social factors that are external and that determine content (Burge 1979: 79; 1988: 650), and Hilary Putnam, whose argument for externalism shall be examined in detailed in the following section, has argued that it is causal relations to one's physical environment that determines the meaning of one's terms.

It could also be suggestive to mention Burge's reason for a different label than "externalism". He prefers to calls his position "anti-individualism", because in his view, the term "externalism" "invites a conflation of the locus and properties of the mental states and events with the locus and character of the environmental relations on which they are constitutively dependent" (Burge 2003: 435). It is true that the label might invite a conflation, and this is why some considered it necessary to distinguish between strong and weak version of externalism.

For example, according to Cynthia MacDonald, a strong version of externalism is a claim about mental state instances ("tokens") and their location (their being "outside the subject of those states"), and it implies that mental state tokens are located, in part, outside the subject of those states (MacDonald 1989). However, according to Mark Rowlands, both the strong form and a weak form of externalism make a claim about mental state *types*, yet from here he agrees with MacDonald that, according to the strong version, "mental states are constituted by or composed of environmental objects or properties" (Rowlands 1995: 360). The reason why Burge prefers a different label is because, in his view, externalism should not be conflated with what gets called a "strong version of externalism" (no matter whether it is a claim about types or tokens). As Burge puts it, "the congenially loose talk, which derives from Putnam's original paper, about what is and what is not in the head should be laid aside" because "spatial location is not the central issue" (Burge 2003: 435-436). What *is* at issue is the question as to what determines content, or, as Burge puts it, the question about the factors on which having particular mental properties (having of content, meaning, etc.) "constitutively depend".

Recall Dretske's confession that he was unable to conceive of a materialistic theory of psychosemantics – a theory which, among other things, aims to answer the question of what makes A's thought about *x* or have content *p* – that isn't externalist in character. And it should be clearer by now as to what might be the implicit premises due to which he cannot conceive of one, as well as be clearer as to what it could mean for a theory to be externalist in the relevant sense.

Of course, being committed to the scientific form of semantic naturalism, a naturalistic theory of semantics ought to include only those "external" factors which satisfy the constraints of scientific naturalism. For example, Burge's reference to *social* factors could not meet the constraint of a naturalistic theory being reductive, and whether it can be called "naturalistic" depends on whether one can develop a form of non-reductive naturalism. One way or the other, both – semantic naturalism as well as semantic externalism – are part of what Glüer and Glock call a *philosophical meaning theory* and Stalnaker calls *foundational semantics*, and it might be useful to get clearer as to what being a theory of this kind might amount to.

According to Glüer, "in the philosophical theory of meaning a distinction can be drawn between theories aiming at correct semantic description of individual languages and theories the subject of which is linguistic meaning as such" (Glüer 2014: 84). Theories of the first kind concern both formal languages such as first-order predicate logic as well as natural languages such as Lithuanian or Yiddish. They are also "usually formally worked out and, therefore, the discipline of developing them is called 'formal semantics'" (ibid.). By contrast to its formal counterpart, a *philosophical* meaning theory is, as Glüer puts it, a more foundational enterprise, for it concerns the nature of meaning or content itself. In her view, it "concerns the place of semantic facts in a wider metaphysical space" (Glüer 2014: 84).

When drawing what is essentially the same or very similar distinction, Glock writes as follows:

> A theory of meaning in the new formal vein does not explain *what it is* for an expression to be meaningful, or to have a specific meaning. Instead, it generates for each actual and potential sentence a particular language *L* a theorem that specifies the meaning of that sentence. […] The new theories aim at *constructing a calculus*, which will enable us to compute the meaning of a complex expression from the meaning or meaning-relevant properties of its simple components. (Glock 2012: 52)

So according to Glock, there are formal theories the desideratum of which is "to provide an explanation for a range of linguistic phenomena such as semantic compositionality and productivity, intensionality, synonymy, ambiguity, reference, truth-aptness, inferential relations, etc." (Ibid.). However, in contrast to the formal theories, there are philosophical theories of meaning, that aim to explain *what it is* for an expression to be meaningful – in other words, theories that aim at a constitutive account of meaning or content.

Again, what seems to the essentially the same distinction is also drawn by Stalnaker. He writes as follows:

> First, there are questions of what I will call 'descriptive semantics'. A descriptive semantic theory is a theory that says what the semantics for the language is, without saying what it is about the practice of using that language that explains why that semantics is the right one. (...) Second, there are questions, which I will call questions of 'foundational semantics', […] about what makes it the case that the language spoken by a particular individual or community has a particular descriptive semantics. (Stalnaker 1997: 535)

If Stalnaker's *what makes it the case* that X is meaningful is the same question as Glock's question as to *what it is* for X to be meaningful, and if both of them are the questions that concern the place of semantic facts in a wider metaphysical space, then the distinction is, ultimately, between *formal* vs *philosophical* semantics.

Finding a location for naturalistic theories of semantics and semantic externalism is relatively straightforward, given how Glüer describes what has

been one of the main interests of current philosophical meaning theories. She writes:

> In the second half of the Twentieth Century, philosophers of language have been especially interested in the relation between semantic facts and facts that can be described in naturalistic terms, and different versions of reductive and non-reductive naturalism have been discussed. Another, though related, debate concerns the question whether the facts determining meaning (and thought content) are facts in some sense internal, or external, to the subject saying or thinking something. (Glüer 2014: 84)

It should be clear by now that naturalistic theories of semantics that are committed to scientific naturalism – the standard view – must be both reductive and externalist.

As Dretske pointed out in the passage quoted at the beginning of the section, in particular cases, the relevant natural and "external" factors might differ. For example, according to informational semantics, content is determined by nomic (i.e. lawful) relations that internal physical states bear to items in the external world. In a simple version of this theory, content is determined by reliable correlations between internal physical states and the items in the world that cause them. In a more sophisticated version, content is determined by an internal physical state's "subjunctive career" (Fodor 1990: 58): in essence, the idea is that it is not enough that the internal physical state is caused by X; an internal state has content X only if that state wouldn't occur if X were not present. It is clear that reference to laws and counterfactuals amounts to a theory being externalist in the relevant sense, and so one need not refer to causation explicitly in order for a theory to be of externalist character.

For example, according to biosemantics, facts about contents are determined by the evolutionary history of the structures of the brain whose contents they are. It too is an externalist theory of content determination, for it amounts to the view that possession of content is determined (at least in part) by factors that are external to the agent. However, as Bridges notes,

biosemanticists' claim that the structures of the brain were *in fact* evolving in the way they say they did "is no less speculative than the hypotheses advanced by evolutionary psychologists about the selectional pressures producing various social and interpersonal phenomena" (Bridges 2006: 524). Hence, whether the account of content determination that biosemantics provides is true also depends on whether the class of such facts is not empty.

So all of the above theories are externalist in the sense that they appeal to factors which are external to the subject, are most likely unknown to the subject (the subject is not aware of these factors), and which determine the contents and objects of thought. But commitment to externalism is not sufficient for a constitutive theory of intentional attitudes to be a bifurcationist one, and it certainly does not imply Cartesianism. In order for a theory which is externalist in character to manifest a Cartesian and bifurcationist picture of mind, it must also be the case that it has some specific theoretical commitments as to the nature of what our mental concepts apply to. More specifically, it must be committed to the view that what our mental concepts pick out needs to be related to the external world (thus justifying the necessity of the "external" element) for the intentionality of mental states to be possible.

However, before examining two of the most predominant philosophical theories of mind and their implications with regards to the subject matter of mental concepts, it might be suggestive to provide an analysis of Putnam's argument for semantic externalism. For as McDowell notes, "it is widely supposed that Putnam's considerations compel a "duplex" conception of at least large tracts of our thought and talk about the mental" (McDowell 1998a: 278). So the analysis of Putnam's considerations on mind might not only be useful for a better understanding of what externalism amounts to. It also can contribute to understanding the reason why one might feel compelled to the bifurcationist conception of the mind, as well as to the clarification of what could it mean.

3. 2. Putnam's externalism and the Cartesian target

One of the most influential arguments for semantic externalism has been presented by Hilary Putnam in his "Meaning and Reference" (1973), and the subsequent paper "The Meaning of 'Meaning'" (1975a) which he considered to be "a very much expanded version" (Putnam 1973: 699) of the previous paper. In his view, the primary goal of these papers is to clarify the ordinary notion of meaning. In Putnam's words,

> analysis of the deep structure of linguistic forms gives us an incomparably more powerful description of the syntax of natural languages than we have ever had before. But the dimension of language associated with the word "meaning" is, in spite of the usual spate of heroic if misguided attempts, as much in the dark as it ever was. (Putnam 1975a: 131)

The way Putnam aims to contribute to the clarification of the notion is by means of a set of arguments that are based on a series of thought experiments. Differences aside, all of the proposed thought experiments have the following pattern: Putnam invites the reader to imagine a counterfactual individual who is exactly like some actual individual with respect to "purely internal" *psychological* and *physical* properties, but who is situated in a counterfactual environment which differs from the actual one in some subtle way. In the best known case, Putnam invites us to imagine that in the year 1750 there was a planet which he calls "Twin Earth" and which was exactly like our Earth. So everything is exactly alike except that instead of water, which Putnam assumes to be H_2O, Twin Earth contains a substance whose chemical composition is different – it is XYZ: the liquid, the reader is told, has a different "hidden structure" (Putnam 1975a: 235, 241; 1986: 288; 1988: 36). In Putnam's story, in 1750 nobody could distinguish between water and twin-water, or twater. Yet, according to Putnam, an individual on Earth using the term "water" in 1750 would be referring to H_2O and not to XYZ; conversely, if the term were used on Twin Earth, it would refer to XYZ, and not to H_2O. Given that

94

everything "in the head" is the same, and assuming that reference is determined by meaning, what the term "water" means is not determined by what is "in the heads" of those individuals. Hence, Putnam concludes, "'meanings' just ain't in the head" (Putnam 1975a: 144).

It should be noted that Putnam's thought experiments were primarily used to establish externalism with respect to natural kind terms such as "water" or "aluminum", so in its original form a thesis is about linguistic meaning of a specific type of linguistic expressions. For this reason, Putnam's externalism is sometimes called "natural kind externalism". However, soon after "The Meaning of 'Meaning'" was published, Tyler Burge and Colin McGinn both argued that what Putnam claims to be true of linguistic meaning of natural kind terms, is also true of the contents of propositional attitudes more generally, and so the scope of externalism should not be restricted to some narrow range of terms (Burge 1986, 1988; McGinn 1989).

In order to get the full force of the thought experiment, one must consider it in conjunction with Putnam's aim to examine the "two unchallenged assumptions" which are, in his view, implicit in the theories of meaning defended by Gottlob Frege and Rudolf Carnap. Here are the two assumptions:

> (I) That knowing the meaning of a term is just a matter of being in a certain psychological state (in the sense of "psychological state," in which states of memory and belief are "psychological states"; no one thought that knowing the meaning of a word was a continuous state of consciousness, of course).
>
> (II) That the meaning of a term determines its extension (in the sense that sameness of intension entails sameness of extension). (Putnam 1975a: 135-136)

It is of crucial importance here to take note of the fact that by the notion of "psychological state", as used in (I), Putnam maintains that he does *not* mean "a state which is studied or described by psychology" (ibid.: 136) because in that case (I), in his view, "may be trivially true" (ibid.). Rather, Putnam claims,

when traditional philosophers talked about psychological states (or 'mental' states), they made an assumption which we may call *the assumption of methodological solipsism*. This assumption is *the assumption that no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed*. (Italics – M.G.) (Ibid.)

The idea of methodological solipsism, according to Putnam, "is pretty explicit in Descartes, but it is implicit in just about the whole of traditional philosophical psychology" (ibid.: 136-137). Accordingly, the notion of a "psychological state", as used by Putnam, should be interpreted as referring to the kind of state which is defined by the principles of methodological solipsism[6].

The principal goal of Putnam is to show that assumptions (I) and (II) "are not jointly satisfied by *any* notion, let alone any notion of meaning" (ibid.: 136), which shows, in his view, that "the traditional concept of meaning is a concept which rests on a false theory" (ibid.). In his view, (I) and (II) imply that "two speakers cannot be in the same psychological state in all respects and understand the term *A* differently; the psychological state of the speaker determines the intension (and hence, by assumption (II), the extension) of *A*" (ibid.: 139). However, he takes the implication to be false, and the reason *why* it is false is that, in Putnam's view,

it is possible for two speakers to be in exactly the *same* psychological state (in the narrow sense), even though the extension of the term *A* in the idiolect of the one is different from the extension of the term *A* in the idiolect of the other. Extension is *not* determined by psychological state. (Ibid.)

---

[6] It is worth noting that the idea of methodological solipsism, as defined by Putnam, is generally taken to be expressing one of the defining features of psychological internalism. For example, according to Susana Nuccetelli, the doctrine of internalism can be defined by reference to two ideas: "***Int*** Mental properties are local properties of individuals.
InP₁ A property is local, internal, or intrinsic if and only if it does not presuppose the existence of anything other than the contingent object that has it." (Nuccetelli 2003: 2).

Putnam claims that there are two ways to counter the conclusion: "to give up the idea that psychological state (in the narrow sense) determines intension, or to give up the idea that intension determines extension" (ibid.). After considering both of them, he decides to give up the first: hence, the property of meaning something is *not* determined by a *narrow psychological state*. Rather, in Putnam's view, it is determined by causal interactions with the natural kind itself, which is a manifestation of causal theory of reference.

The argument can be reconstructed in the following way:

(1)	Psychological states are narrow.	*Methodological solipsism*
(2)	It is possible to be in the same narrow state yet refer to different objects.	*Twin Earth thought experiment*
(3)	Narrow states do not determine reference.	*From (2)*
(4)	Reference is determined by meaning.	*Assumption*
(5)	Narrow states do not determine meaning (or reference).	*From (3) & (4)*
(6)	Psychological states do not determine meaning (or reference).	*From (1) & (5)*
(7)	Meaning (and reference) is determined by causal interactions with the relevant natural kinds.	*Causal theory of reference*

As noted by Phil Hutchinson, "when Putnam writes 'meanings ain't in the head' he takes himself to be meaning that they 'ain't in' a *person's* 'mind'; they are external" (Hutchinson 2008: 22), though this should be taken with care. For the idea that they are "external", and not "in a person's 'mind'" is not meant to be an idea about their *location* – recall Burge's point and reasons for a different label, explicated in the previous section. Rather, what is meant is that *what some term means* is not determined by a *narrow psychological state*, but rather by factors that are external to the mind *thus conceived*. Of course, the positive theory as to what determined content does mention factors which are external to an individual. For according to Putnam, (i) it is what is in the

extension of some natural kind term that determines *what* the term means, and (ii) it is causal relations (although, presumably, in conjunction with the "narrow" psychological states) which determine *that* some term has the meaning (and reference) that it does. Hence, meaning, according to Putnam, is at least in part determined by external factors (in the sense of (i) and (ii)) of which the subject need not be aware.

Putnam's considerations about Twin-Earth are generally taken to show that content externalism is true. However, its implicit commitment to the doctrine of methodological solipsism is rarely made explicit, although it is of conceptual importance. For one, it makes it explicit that the psychological state which was supposed to play an important role in arguing for the view that being in it is insufficient for determination of what the term means, is narrow. Secondly, and relatedly, it might explicate the reason as to why one could then think that it must be something external to the mind *thus conceived* that determined meaning or reference. It seems that at least part of the reason for bringing in external factors in determining meaning or reference is an implicit commitment to the idea that psychology is narrow.

It is hard to tell whether Putnam's considerations about meaning imply a bifurcationist account of the mind. Perhaps they don't imply it. But it is not hard to imagine how these considerations might compel one to adopt a duplex picture of the mind, for they presuppose all the necessary and sufficient materials for it to be one. If a theory of intentional mental states is then construed on the basis of these materials – that is, if the relevant mental states are analyzed in terms of narrow states and the external factors which determine the meaning and reference of terms – then the theory does seem to end up being an essentially bifurcationist picture of mind.

Recall that semantic naturalists aim to provide a constitutive theory of intentionality – a theory which tells what it is for a creature to be in some mental states with particular content. Thus, by assuming semantic externalism they are therefore committed to the view that external factors are part of a constitutive theory of intentionality. This, in turn, implies that semantic

naturalists satisfy one of the two conditions for a theory to be of a bifurcationist kind – they agree that an external element is constitutive for thought, so it is necessary in the analysis of intentionality of thought. But as it has already been noted (section 3.1.) this is not sufficient for bifurcationism, because a bifurcationist picture of the mind requires there being two factors – an external *as well as internal* element – both of which are necessary and jointly sufficient for the phenomena that is being analyzed. So the remaining question then is whether the general framework of semantic naturalism contains another (internal) factor, that too is considered to be constitutive for thought. More specifically, this turns into a question whether the general framework on which the predominant versions of semantic naturalism rest implies that mental concepts apply to something "wholly in the head" which, in turn, needs to be related (by the external element) to the world outside the head for intentionality of thought to be possible.

At this point, it should be made explicit that an inquiry into whether semantic naturalism is committed to something being "in the head", or to something that is "internal" in some sense, rests on unavoidable yet not radical semantic indeterminacy. For one, not only one cannot know in advance the conceptual framework that is implicit in the theory, but neither can one decide in advance what it is *exactly* that one is looking for in this case. Part of the reason why an inquiry into the "internal" or the Cartesian elements of mind is not clear-cut is well captured by Robert Stalnaker. He writes as follows:

> Wittgenstein, Ryle, Quine, Sellars, Davidson (not to mention Heidegger) may have cut off a few Cartesian heads, but they keep growing back. Descartes is not the bogeyman he once was; Cartesian skeptical arguments, and arguments for the autonomy of minds and mental states are back in fashion, and philosophers feel free again to observe and contemplate the inner objects that Wittgenstein tried to banish. (Stalnaker 2008: 1)

As Stalnaker aptly points out, the main reason why the Cartesian picture of the mind continues to color our philosophical projects is that "the Cartesian target

is of course a broad and diverse one: critics of one aspect of the picture may embrace another" (ibid.: 2). For example, Davidson considered Quine's naturalized epistemology to be "Cartesian in spirit and consequence" (Davidson 2005: 58), and rejected "Quine's account of the nature of knowledge, which is essentially first person and Cartesian" (Davidson 1991: 192).

Given that the categories of "being in the head", "being internal", "being part of the inner realm" or "being Cartesian" are not clear-cut and clearly do not have sharp boundaries, it is better to start in the middle: that is, to start with a general idea that gets abstracted from many agreed cases of being Cartesian, or being wholly in the head, or being part of the inner realm, and proceed with that. The main goal of the following two chapters is to explicate some of the general philosophical presuppositions that constitute the orthodox position in contemporary philosophy of mind, and decide whether the presuppositions commit one to the idea that our mental concepts apply to the items of a kind, which need to be related to the world outside the head for the intentionality of thought to be possible. If *that* can be shown, then it should be sufficient to demonstrate that a form of Cartesianism is implicit in the picture, for it would show that a theory is committed to there being something that can reasonably considered to be "wholly in the head", "purely mental" or not, and which is what our mental concepts are supposed to pick out. If so, then there seems to be good reason to believe that naturalistic theories of semantics which are externalist in character and do rest on these philosophical presuppositions manifest what appears to be a Cartesian and bifurcationist picture of the mind.

4. Functionalism in the philosophy of mind

4. 1. The origins of functionalism

Functionalism as a philosophical theory about the nature of mentality has developed out of the critical reflections on its two predecessors: the doctrine of

behaviorism, with its psychological and philosophical modes, and a philosophical theory that is commonly known under the name of the "identity theory of mind". The doctrine of functionalism is commonly taken to inherit the virtues of its predecessors without being committed to their vices. Hence, in order to attain a more perspicuous representation of functionalism itself, it is of great importance to explore the conceptual frameworks of the theories from which it has originated. The first predecessor of functionalism was behaviorism.

It is common to distinguish between two kinds of behaviorism: *philosophical behaviorism* and *psychological behaviorism*. According to Rowland Stout,

> philosophical behaviourism is a view about the nature of the mind, the concept of mind and mental predicates. It may remain neutral about how the science of the mind should be pursued. Psychological behaviourism on the other hand is committed to the methodological claim that the scientific study of animal psychology should be limited to the scientific study of animal behaviour. (Stout 2006: 21)

Stout is right about the main difference between the philosophical and the psychological kind of behaviorism: namely, the idea that the former may (and according to some – should) remain neutral with regards to empirical psychology, or the science of the mind. But despite the indicated difference, there are two important interrelations between the two kind of theories that should be made explicit.

First, both philosophical and psychological behaviorism is committed to what can be called a principle of intersubjective verifiability, which has been given different interpretations in the domains of the philosophy of language and the philosophy of science, the latter being relevant for the psychological, the former – to the philosophical type of behaviorism.

In the philosophy of science, the principle expresses one of the main epistemological axioms of empirical investigation and scientific inquiry: the idea that the domain of any empirical inquiry must be accessible and verifiable

(in principle, if not in fact) intersubjectively. However, not always has this epistemological principle been guiding the scientific inquiries of psychology. In the XIX century, when psychology was primarily focused on what was then considered to be its traditional domain – the domain of consciousness – its epistemology was based on the principle of introspection. And it was so until J. B. Watson expressed his distrust in the value of introspectively acquired evidence. According to Watson, if the science of psychology aims to be an objective science, its experimentations shall begin and end with the study of behavior. In his words, "psychology as the behaviorist views it is a purely objective experimental branch of natural science" (Watson 1913: 158).

When applied to the domain of the philosophy of language, the principle becomes the primary guide for a *reductive* type of philosophical behaviorism that was developed by logical positivists. Many different formulations of the verifiability principle have been proposed, but the main idea lying behind verificationism insists on there being a link – stronger or weaker – between the meaningfulness of empirical statements (i.e. statements that report empirical facts) and their verification conditions. Those who subscribed to the principle, held that the only cognitively meaningful statements are statements that are empirically verifiable, either in fact or in principle. If verifiability cannot be achieved, then statements are either logically necessary (and so do not state matters of fact) or meaningless.

In the case of theoretical statements, logical positivists subscribed to a version of semantic foundationalism and held that all theoretical statements of empirical psychology draw their meaning from statements about physiology and behavior, i.e. propositions concerning the alleged semantic foundation of the theoretical vocabulary of psychology. When the principles of semantic verificationism and foundationalism are applied to empirical psychology, they serve as the semantic ground for the reductive form of philosophical (or logical) behaviorism.

At this point it shall be noted that the term "philosophical behaviorism" is not only used to apply to a reductive form of logical behaviorism as it was

defended by the members of the Vienna Circle. For one of the core principles of philosophical behaviorism – namely, that mental concepts are to be given a dispositional analysis – has also been defended by philosophers who neither subscribed to the verifiability principle, nor accepted the idea of semantic foundationalism. Moreover, philosophical behaviorists such as Malcolm, Ryle or Wittgenstein were neither interested in proposing semantic analyses for theoretical statements of empirical sciences, nor were they inspired by the reductive ambitions of the logical positivists more generally. Their primary target was ordinary, not scientific language, and their primary aim was conceptual elucidation of ordinary psychological concepts, not reformation that would fit certain philosophically grounded scientific ideals.

The second similarity between the two types of behaviorists is that they unanimously reject the idea that mentality is something *internal* to an individual. According to behaviorism, there are no internal entities (events, states or processes) that are mental, yet the idea is perfectly compatible with the view that there are events, states or processes of a different kind. Rather, mentality, in their view, must be open to public observation and verification, as the principle of intersubjective verifiability requires, and the principle is satisfied by patterns of animal behavior.

However, it is important to note here that the conception of behavior implicit in the frameworks of behavioral psychologists such as Watson and Skinner, as well as in logical positivism, was based on proximal stimulus and proximal response patterns. It is this kind of behavior that the reductive type of behaviorists called "dispositions", and it is *this* conception of dispositions that has prevailed and dominates the debates in the contemporary philosophy on the nature of dispositions or the nature of action. The stimulus-response conception of animal behavior has been by and large inspired by the experimental works on reflex reactions and classical conditioning that were performed by Russian physiologists Ivan Pavlov and Ivan Sechenov. According to this conception, animal behavior is the response (the 'proximal output') of the stimulus that is considered to be the 'proximal input' (occurring

at the skin of the animal) which is coming from the external environment, both of them being observable phenomena. The stimulus-response patterns can be altered either via conditioning, or through associative learning. In so-called classical conditioning, a neutral stimulus is followed by an unconditioned stimulus which induces a reaction; in "operant conditioning", the presentation of the stimulus depends contingently on a particular behavioral response to an initially neutral stimulus. But neither classical nor operant conditioning was supposed to involve genuinely mental phenomena: the idea that there are mental events, states or processes was either denied or disregarded on methodological grounds as falling outside the scope of scientific methodology. In the first case, one arrives at a position that can be described as *ontological* behaviorism; in the latter case, one arrives at *methodological* behaviorism. The complex and apparently intelligent animal behavior is explained by the behaviorist purely in terms of stimulus-response mechanisms, whether innate or acquired. Genuinely mental and in particular cognitive processes and capacities are either ignored or denied.

The influence of behaviorism in psychology diminished with the rise of the so-called cognitive revolution, or cognitivism, in the 1960s. M. J. Cain describes the gradual yet radical shift in the following way:

> From the early years of the twentieth century into the 1960s, behaviourism constituted the dominant approach in scientific psychology in the English-speaking world. With the birth and development of cognitive psychology and cognitive science in the 1960s, this behaviourist dominance was challenged and behaviourism gradually fell into disrepute. As a result of this 'cognitive revolution' psychologists came to operate with a quite different conception of the research agenda of their discipline. They came to see their central concern as being that of explaining intentionally characterized cognitive capacities and held that in order to explain such capacities, it is necessary to appeal to internal representational states and processes. (Cain 2002: 20-21)

Stout gives a similar description:

> While accepting a broadly behaviourist denial of introspection, cognitive psychology rejected the behaviourist claim that the subject matter of psychology is just patterns of behaviour. Cognitive psychology looked for mechanisms behind these patterns and found them by positing internal representations as causally explanatory entities. (Stout 2006: 37)

So, since the cognitive turn, cognitive psychologists no longer hold that the subject matter of empirical psychology consists in behavioral patterns. Neither do they subscribe to the stimulus-response conception of animal behavior or behavioral dispositions. Rather, after the cognitive revolution the primary goal of empirical psychology is to explain "intentionally characterized cognitive capacities", and it is widely believed that this must be done by means of positing "internal representational states and processes". The revolution was partly initiated by Chomsky's (1959) reaction to Skinner's behaviorist linguistics. But an important impulse was given by the progress of computer science, which seemed to open the perspective of a genuinely mechanistic explanation of cognition.

However, it is not only a different *conception* of the explanandum – from stimulus-response patterns to intentionally characterized cognitive capacities – that has been adopted after the cognitive revolution. Unlike their predecessors, the cognitivists have also attained a different conception of behavior as well as a different view of its explanation. As Glock notes,

> cognitive psychology was based on two fundamental tenets. First, in order to explain behaviour one must peek inside the black box and posit internal states and episodes, notably beliefs and desires. Secondly, these inner phenomena are 'mental representations', cognitive intermediaries between perceptual input and behavioural output. (Glock *forthcoming*: Ch. 1)

Glock is right in maintaining that cognitive psychologists, unlike the behaviorists (except Ryleans or Wittgensteinians), did not conceive of behavior

in terms of stimulus-response patterns; rather, they held that behavior involves mental phenomena and that these phenomena are what constitute the explanandum of cognitive psychology. But the idea that mental states (beliefs, desires, etc.) are mental representations which are posited to explain cognitive capacities does not follow from the cognitivist turn itself. Rather, it is a consequence that follows from the conjunction of empirical theorizing and ontological commitments of cognitive psychology with the philosophical doctrines of functionalism and the representational theory of mind. For the view that mental representations are posited to explain cognitive capacities does not, by itself, imply that mental states *are* mental representations, unless one assumes so on independent grounds.

There are two types of arguments that are offered against philosophical behaviorism. As noted by Block, "probably the most influential argument against behaviorism is due to Chisholm and Geach" (Block 1981: 11-12), which is supposed to show that "one cannot define the conditions under which a given mental state will issue in a given behavioral disposition without adverting to *other mental states*" (ibid.). The so-called intentional circle has since been considered as one of the main weaknesses of behaviorism, and the ability to put it at rest is considered to be one of the main positive aspects of functionalism. However, it shall be noted that the intentional circle argument, which is unanimously taken to be an argument against philosophical behaviorism *per se*, has force only if directed against the *reductive* type of behaviorism defended by logical positivists. As already noted, neither Cambridge, nor Oxford behaviorists aimed at providing reductive analyses of mental concepts, and Ryle accommodated the intentional circle into his analysis of mental phenomena.

The second type of argument that is commonly directed against philosophical behaviorism targets the idea that is also implicit in psychological behaviorism. As noted by Lycan, philosophical behaviorism was becoming less stringent once philosophers "felt that in its total repudiation of the inner, Behaviorism was leaving out something real and important" (Lycan 2003: 49).

According to Barry Dainton and Howard Robinson, "resistance to behavioristic approach did not come from within the philosophy of science, but from the intuitive feeling that any purely behavioral approach left something out" (Dainton & Robinson 2014: 131). For example, U. T. Place and J. J. C. Smart held that the dispositional analysis of what Place called "cognitive concepts" (such as knowing, believing, understanding, remembering) is "fundamentally sound", yet with regards to concepts "clustering around the notions of consciousness, experience, sensation, and mental imagery", he believed that "some sort of inner process story is unavoidable" (Place 1956: 44).

Behaviorists' banishment of the internal was considered to be a mistake for at least two reasons. The first is that the internal states are not only considered to be internal in the sense of being literally inside the individual; they are also believed to be essentially private, and it is *this* feature that is commonly held to be necessary to explain the possibility of subjective character of some of the mental states, including intentional states. The second reason stems from considerations regarding mental causation – more specifically, the idea that mental states are internal is commonly taken to be a necessary condition for any adequate account of mental causation, which behaviorists did not meet. Thus, the defining qualities of physicalism, or materialism, were to be modified, and considerations regarding the nature of consciousness on the one hand, and the nature of action and mental causation on the other were features that "led U. T. Place, J. J. C. Smart, and D. M. Armstrong to augment their behavioral approach with central state materialism, identifying the mental states not with the rather abstract entity, a *disposition* to behave, but with the solid inner machinery that caused the behavior" (Dainton & Robinson 2014: 131). This has naturally led to yet another predecessor of functionalism – the identity theory of mind.

In the philosophical literature, the mind-body identity theory is commonly introduced in one of the two ways: it is characterized either as a theory which preceded and has been rejected by the doctrine of functionalism

(e.g. Putnam 1967), or as a theory which is entirely consonant, indifferent to (e.g. Smart 1959), or even entailed by (e.g. Lewis 1972) its successor.

Despite important differences between the two kinds of theories, proponents of both types stress the theoretical importance of two general ideas that are supposed to make their doctrines different from the preceding philosophical theories of mind. The first idea that underpinned the identity theory was that it neither presupposes nor entails synonymy of mental and physical predicates. Thus, according to one of its main proponents,

> "Consciousness is a process in the brain" […] is neither self-contradictory nor self-evident; it is a reasonable scientific hypothesis, in the way that the statement "Lightning is a motion of electric charges" is a reasonable scientific hypothesis. (Place 1956: 45)

The proposition that statements like "consciousness is a process in the brain" are neither self-contradictory nor self-evident is supposed to follow from the idea that such claims do not purport to explicate the meanings of mental terms. In Fregean terminology, identity statements are meant to state identity of reference and not identity of sense.

The second idea is that the proposed identities should be understood as expressing "reasonable scientific hypotheses". As Place notes, "the thesis that consciousness is a process in the brain is put forward as a reasonable scientific hypothesis, not to be dismissed on logical grounds alone" (ibid.: 44).

The relevant differences between the two kinds of identity theories consist in different interpretations as to what the proposed identities are identities *of*: not as to whether the identity is an identity of reference, as opposed to the identity of sense, but as to *what it is* that the co-referential notions signify.

One kind of identity theory is described by Frank Jackson. He writes:

> identity theorists appeal to scientific identities in explaining and introducing their theory. […] When scientists tell us that lightning is an electrical discharge, they are not merely telling us that the instance or

token of lightning we saw last night is an instance or token of an electrical discharge; they are telling us, in addition, about *what kind of happening lightning in general is*. […] This suggests that we should think of the (mind-brain) identity theory as a type-type identity theory. Moreover, the theorists' favourite illustration – 'Pain = C-fibres firing' – is a type-type identity statement. (Italics – M.G.) (Jackson 1998: 398)

The above passage indicates that one type of identity theory is a *type* or *property* identity theory which commonly goes under the name of "type physicalism". Consequently, the identity of *reference* that is being expressed by the identity statements is an identity of *types* or *properties*. But if the identity theory is understood in this way, then three points should be made clear in order to understand how identity theory of this kind differs from its counterpart.

First, if identity statements proposed by the identity theory are supposed to tell one what kind of thing some mental property "in general is", then suchlike identity statements are of the same kind as identity statements like knowledge is degettierized justified true belief or colors are dispositions to produce experience. Neither of them tells a story which is only about the instances of knowledge or colors, and both of them tell us what kind or type of thing knowledge in general is, or what kind of thing a statue is.

Second, the identity statements which aim to tell what kind or type of thing something is do not assert identities of objects that are approached from two different epistemological standpoints. So it is not as if one can either observe, or think of something as *knowledge* via one epistemological or semantic mode of presentation, and as *degettierized justified true belief* via another mode of presentation; or as *action* from one epistemological or semantic perspective, and *a bodily movement caused by some mental state* from another. In such cases of identity, there is no object or entity that could have two incompatible descriptions, as there is in cases of identity statements like "Cicero = Tully" or "Hesperus = Phosphorus".

Third, in cases where identity statements are meant to express identities of properties or types, they have the following implication: for any properties P and Q, if P = Q, then *a* has P if and only if *a* has Q. Take the case of color. If the properties of being navy or being grey *are* dispositions to produce experience of seeing navy or grey, as some philosophers hold, then *a* is navy or grey if and only if *a* has a disposition to produce the relevant kind of experience. It follows from this that if for any mental predicate MP there is some physical predicate PP such that MP = PP, then *a* has MP if and only if *a* has PP. So if the property of believing that p is identical to the property of, say, D-fibers firing, then S believes that p if and only if S's brain fires D-fibers. The equivalence follows from, and is explained by, the identity of properties – the objects of reference in this case – that are signified by the two predicates.

When functionalism is described as a doctrine that supersedes and rejects the identity theory, the kind of identity theory that is being rejected is *type* or *property* identity theory. This is explicit in Fodor:

> It looked, in the early 1960s, as though anybody who wanted psychology to be compatible with a physicalistic ontology had a choice between some other kind of *behaviourism* and some or other kind of *property-identity theory*. For a variety of reasons, neither of these options seemed very satisfactory (in fact, they still don't) so a small tempest brewed in the philosophical teapot. (Fodor 1985: 81)

One of the main reasons why type- or property-identity theory is widely considered to be inadequate as philosophical theory of mentality is because it is "a *chauvinist* theory: it withholds mental properties from systems that in fact have them. In saying mental states are brain states, for example, physicalists unfairly exclude those poor brainless creatures who nonetheless have minds" (Block 1978: 270). Thus, type identity theory is widely considered to be incompatible with such neurophysiological facts as variable realization and neural plasticity: empirical facts that any defensible philosophical theory of mentality must be able to accommodate.

It is commonly held that there are two options for a philosophical account of mentality to be made compatible with a physicalistic ontology. As Jackson notes,

> identity theorists can retreat to a token-token identity theory. Each and every token or instance of mental state M is some token brain state, but mental types are not brain types, being instead functional types. Alternatively, they can allow that identities between mental types and brain types may need to be restricted. (Jackson 1998: 389-390)

But even those who endorse the second option, such as Jackson himself, generally admit that the restricted type identity theory cannot offer a general account of what some mental type is. Consequently, anyone willing to provide a general theory of mentality must retreat to functionalism with regards to mental types that would be accompanied by a token-identity theory. This, in turn, explains the fact that the identity theory which is supposed to *follow* from the doctrine of functionalism is a kind of identity theory according to which the identity of reference is not an identity of properties or types, but identity of their instances or "tokens". Identities of *this* kind, however, are not to be considered as expressing truths that tell what kind of thing some mental property in general is, so they are not meant to express identities of the type which were proposed by its early defenders. And it is at this point where the doctrine of functionalism plays an essential role in giving a constitutive account of the nature of such states.

4. 2. Functionalism and the idea of internal states

Unlike the type of identity statements that underlie type physicalism, the identity statements which are put forward by functionalists or token identity theorists are commonly considered to be on a par with identity statements that express identities between *objects*: so identities like "Hesperus = Phosphorus" or "Aphla = Ateb". Contrary to the type of identity statements that underlie

type physicalism, identity statements like "Hesperus = Phosphorus" *do* assert identities of objects and in that respect differ from identity statements like "knowledge = degettierized justified true belief" or "colors = dispositions to produce experience etc.". For identities of the former type *do* rest on the idea that one and the same object can be observed, or thought of, from two different epistemological or semantic standpoints. Moreover, they are meant to be empirically informative: observation can disclose that what was held to be two distinct objects is in fact one, or vice versa.

To avoid the logical consequences of type physicalism, the new kind of identity theory has to provide an alternative conception of the object of reference of the identity statements. And a version of the identity theory that is hospitable to the doctrine of functionalism maintains that the identity of reference is an identity of *states* as opposed to an identity of *types* or *properties*. States are *instances* of types, but this does not imply that state identity can obtain only if instantiated types are identical too. In fact, if identity theory aims to maintain identity of states without being committed to identity of types, it must allow state identity to be compatible with difference in types: it must allow, that is, that exemplifications or instances of mental properties can be identical with exemplifications or instances of physical properties *without* it being the case that the properties *themselves* are identical. Such a requirement can be met if the idea of an "identity of states" is to be understood on the model of identities like "Hesperus = Phosphorus", for it seems that it is only in such a case that identity of states *does not* imply identity of properties (the fact that Hesperus is Phosphorus does not imply that the property of *being the evening star* is identical to the property of *being the morning star*).

But identity theories that argue for the identities of instances or exemplifications of properties do not explain the general nature of the property being instantiated. For example, an identity statement like "Tully = Cicero" does not explain (and does not aim to) what it is for something to be Tully or be Cicero. In that respect, a token identity theory does not – because it cannot –

provide a general account of the mind. This, in turn, is to be done by the doctrine of functionalism.

According to Block,

> "functionalist" theories are the products of a number of rather different projects: attempts to reformulate logical behaviourism to avoid objections, attempts to exploit mind-machine analogies, attempts to apply empirical psychology to philosophy of mind, and attempts to argue for – or against – mental-neurological identity theses. (Block 1978: 261)

In the words of Janet Levin, at its core, "functionalism is the doctrine that *what makes something* a thought, desire, pain (or any other type of mental state) depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part" (italics – M.G.) (Levin 2013: §1).

Thus, contrary to type identity theory yet in perfect compatibility with the doctrine of token or state identity, philosophical functionalism maintains that the nature of mentality is to be characterized by reference to the concept of functional role that a *state* plays in the cognitive system. But at this point it is important to bear in mind that the "something" in Levin's phrase "what makes something…", already has a value when it is interpreted against the background of token identity theory. The *something* is the *object* of identity – the token state – that needs to be made into a thought, desire or pain by a specification of features which make it mental.

According to functionalism, mental properties are second-order functional properties of internal (physical) states. So according to functionalism, for S to V that *p* is for there to be some internal token state with a second-order property which qualifies *it* as being a case of V*ing* (believing, thinking, desiring) that *p*. Accordingly, mental concepts are concepts that apply to internal states of S, and the internal states to which mental concepts apply are states to which S's *mental* states are token-identical: it is a token state which can be picked out both by physical *and* by mental concepts, and so it is a state

which is a kind of particular, that could be conceived of in two semantically and epistemologically distinct ways.

Functionalism is rooted in the causal analysis of mental concepts no less than it is defined by the concepts of function and functional role, and it is of great philosophical importance for the ontology of mind that causal analysis of mental concepts comes with a particular conception of statehood which is part and parcel of the orthodox functionalistic ontology of mind, and which contributes to the particularistic ontology of mental states. The causal analysis is supposed to characterize the nature of states that "play the causal-functional role" in virtue of which states can be considered *as* mental, and it is via the causal analysis of mental concepts that functionalism gets connected to the doctrine of philosophical behaviorism. Like logical behaviorism, the causal analysis of mental concepts is supposed to retain the conceptual connection between mental phenomena and behavior, yet according to functionalism, mental states possess *causal efficacy*: unlike patterns of behavior or behavioral dispositions (at least as they were understood by reductive behaviorists), mental states are supposed to be *causes* of behavior. For example, David M. Armstrong holds that "the concept of a mental state essentially involves and is exhausted by, the concept of a state that is *apt to be the cause of certain effects or apt to be the effect of certain causes*" (Armstrong 1981: 1). Similarly, David Lewis maintains that "the concept of pain, or indeed any other experience or mental state, is the concept of a state that occupies a certain causal role, a state with certain typical causes and effects. It is the concept of a state apt to being caused by certain stimuli and apt for causing certain behaviour" (Lewis 1980: 218). Hence, for functionalism, it is a conceptual-constitutive truth that a mental state is "a state that occupies a certain causal role". As David Papineau puts it,

> the functionalist thinks of mental states as causal intermediaries between
> perceptual inputs and behavioural output. This is an advance on thinking
> of them simply as physical states. But, for all that, functionalism still

presents mental states as part of a system of causal pushes and pulls inside the head. (Papineau 1987: 46)

Given that functionalism is commonly based on a physicalistic ontology, the mental states that are supposed to be a part of "causal pushes and pulls inside the head" are proclaimed to be physical states, more specifically, neural states. Thus, mental concepts apply to neural states of the brain. As Lewis notes, this has an important consequence regarding the identity in question. He writes as follows:

> If pain is identical to a certain neural state, the identity is contingent. Whether it holds is one of the things that varies from one possible world to another. But take care. I do not say that here we have two states, pain and some neural state, that are contingently identical, identical at this world but different at another. Since I'm serious about identity, we have not two states but one. This one state, this neural state which is pain, is not contingently identical to itself. It does not differ from itself at any world. Nothing does. What's true is, rather, that the concept and the name of pain contingently apply to some neural state at this world, but do not apply to it at another. (Lewis 1980: 218)

Thus understood, the doctrine of functionalism implies that some mental state which *is* a neural state – the two states are identical – is characterized as mental (more specifically, as, say, pain or belief) "because" (where "because" is conceptual) the neural state performs a certain causal-functional role. But it is important to bear in mind, that the idea of a state here is to be understood as an idea of something that has causal efficacy, and occupies a certain causal-functional role in the network of causal pushes and pulls inside the head. So a causal analysis of mental concepts is not just an explication of the idea that mental states are causally relevant for behaviour – that believing or being in pain makes a causal difference. Rather, it seems to take for granted a particular conception of what a mental state must be for it to be of a kind that could have the causal relevance that is assumed to have. According to the causal analysis of mental concepts, the state must be a kind of particular which has causal

115

efficacy – it must be *something that causes*. And it is this idea that seems to go hand in hand with the token identity theory, and the idea that internal physical or neural states to which mental states are token-identical are particulars that can have both physical and mental description, and are *causes* that are part of a causal network of "causal pushes and pulls inside the head".

Fodor notes that functionalism thus understood implies

> a new account of the type/token relation for psychological states: psychological-state tokens were to be assigned to psychological-state types *solely* by reference to their causal relations to proximal stimuli ("inputs"), to proximal responses ("outputs"), and to one another. (Fodor 1985: 28)

In Fodor's view, functionalism has two notable advantages over behaviourism and the type-identity theory. One is that it is "*compatible* with physicalism in that it permits tokenings of psychological states to be identical to tokenings of physical states (and thus to enjoy whatever causal properties physical states are supposed to have)"; second, it allows for "tokens of one and the same psychological-state type to differ arbitrarily in their physical kind" (Fodor 1985: 81-82). Hence, the doctrine of functionalism "was greeted with audible joy by the new breed of 'Cognitive Scientists' and has clearly become the received ontological doctrine in that discipline" (ibid.). It not only absorbs the merits and avoids the drawbacks of both behaviorism and type physicalism. More importantly, "if Functionalism is true", as Fodor notes, "there is plausibly a *level of explanation* between common-sense belief/desire psychology, on the one hand, and neurological (circuit-theoretic; generally 'hard-science') explanation on the other" (ibid.). Hence, as Block notes, the doctrine of functionalism allows us to "apply empirical psychology to philosophy of mind" (Block 1978: 261), which, in turn, provides the necessary means to link the ideas of the "cognitive turn" in empirical psychology to the philosophy of mind. Consequently, this results in the representational theory of mind that is predominant in contemporary philosophy of mind.

Before proceeding to the analysis of the representational theory of mind, it is worth pausing to sketch a general picture that seems to be implicit in the

doctrine of functionalism that is rooted in token-identity theory and the causal analysis of mental concepts.

First, if the above analysis of the conceptual framework of functionalism is adequate, then it seems that functionalism is committed to the idea that our mental concepts (concepts like *belief*, *thought*, *intention* or *pain*) apply to an internal token state, a kind of particular, which is supposed to be picked out by its typical causes and effects, and which has some second-order properties that make *it* the mental state (*belief, thought, intentional* or *pain*) it is. As Speaks puts it, in case of *intentional* mental states like *belief, though, intention*, and others, functionalism implies that "the property of believing *p* is constituted by a second-order property of *first-order states of agents* (e.g., the property of being in some state with such and such *relational properties*)" (italics – M.G.) (Speaks 2003: 58). But one must note here that in the context of functionalism, the property of *being in some state* is a property of the brain, for "some state with such and such properties" is a *brain* state. So in fact, the first-order state is a state of the brain, and it is unclear in what sense a state of the brain can be a state of an agent, like a state of her sitting, sleeping or bleeding. In any case, the idea seems to be that the second-order properties are properties of first-order states that an agent's brain is "in". But at this point it is of crucial importance to note that given the conceptual background of token-identity theory the idiom of "first-order state" is supposed to pick out a token state – a kind of particular that is the bearer of second-order properties.

If the above interpretation is adequate, then second-order properties are presumably causal-functional properties of such states. But it is generally agreed that the internal causal-functional properties are insufficient for such states to be about the world; that is, they are insufficient for them to be intentional. To use the common terminology here, the internal causal/functional/computational role of such states at best could only determine their "narrow content" (or "narrow semantics"), which is not sufficient for having *intentional* content. In what way internal causal/functional/computational roles of states of one's brain could determine

any type of *semantics*, is not that clear. But even if one assumes that such roles can in some way or other determine the "narrow content" – as it is commonly held that it can – they would still not determine *intentional* content, because "narrow" content, by general agreement, is not truth-conditional.

The concept of "narrow content" is generally understood on the lines of David Kaplan's notion of "character"[7] that he used in his work on the semantics of demonstratives and indexical expressions[8], and is defined as a kind of function from contexts onto truth-conditional content (see Fodor 1987: 53). But outside Kaplan's context of formal semantics, the relevant "function" here is supposed to be determined by what is "wholly in the head"; more specifically, by the *internal* causal/functional/computational roles of one's internal (physical, brain, neural, whatever) states. However, this cannot be what the functionalists' second-order properties of internal states amount to, for in that case, they would not be sufficient to determine the intentionality of the internal states. Hence, second-order properties of the functionalists' internal states must be determined by states' *external* relations to the world, for it seems that only in that case could possession of second-order ("relational") properties suffice for an internal state that is wholly in the head to have intentional (truth-conditional) content or be about the world (assuming this makes sense). So within the ontological framework of functionalism, it seems natural to hold that it must be something "external" that makes functionalists' internal states intentional.

It might not be what Dretske meant, but given the token-identity based functionalism that is presumably implicit in Dretske's thinking, it seems to be plausible to assume that the "something" in his question as to what "makes

---

[7] The notion has been introduced in Kaplan's "Demonstratives" (1989).

[8] It is worth noting here that it was no part of Kaplan's theory to distinguish a component of semantic content that would be determined by internal causal/functional/computational roles, so Kaplan's *characters* are not "narrow" in the sense required. It seems to be plausible to assume that what Kaplan calls "characters" – the rules that determine the referent of an indexical or demonstrative expression in a particular context – are determined by the *use* of the relevant expressions, and his formal semantics is meant to deliver a formal representation of how the referent (Kaplan's *content*) of those expressions depends on context. Kaplan's formal apparatus does not establish – and it is not meant to – that *character* is determined by something "in the head" whereas its "content" by something "external" to it.

something in my brain about football rather than philosophy" (Dretske 1996: 143) is meant to pick out an internal token state which is a kind of particular that is "in the brain", and whose internal functional, causal, and computational properties are assumed to be insufficient for it to have intentional content or be about the world. So if *that* is what every intentional mental state is token-identical with, and what our intentional mental concepts apply to, then it seems to be reasonable to be motivated to look for an account which would explain how mental states *thus conceived* can be related to the outside world. That is, it seems to be reasonable to assume that if thoughts "are in the head, […] what makes them the thoughts they are is not there", and so to be convinced that there cannot be any "plausible psychosemantics, no plausible theory of what makes one thing about another, that isn't externalist in character", and so no plausible psychosemantics that would not appeal to some kind of relations, "causal, informational, historical, or whatever" (ibid.).

At this point it is important to note that the functionalists' internal token states are not *causally* independent from the external world, because they held to be caused by it and are supposed to have a causal effect on it (yet one should also bear in mind that this is not a sense in which *mental* states are proclaimed to be dependent on the external world by a general characterization of externalism). But on the other hand, neither does there seem to be any good reason to believe that when they *are* caused by whatever they are caused by, it should be sufficient for *them* to be about the world. In fact, it is not clear what could possibly be sufficient for such states to have content or be about something, for there is no clear sense as to what could it *mean* for such states to have intentional content or be about something or other to begin with. Perhaps they are not intentional at all, however caused, even in perfect conditions; they are not about dogs or Napoleon, never mind Pegasus, Zeus or the golden mountain that could not have caused them. But it is not necessary here to decide whether there is any sense in which internal token states could be intentional, and if there is, how does *that* sense relate to the one implicit in our intentional vocabulary. For irrespective of whether such internal states can

be conceived as being intentional or not, the goal of this section was to show that there *is* a sense in which functionalism is committed to internal items of a kind that seem to require them to be related to the things outside the head for intentionality of thought to be possible *on this conception of the mind*. It might not be a "purely Cartesian" one, and in its contemporary mode it is certainly not immaterialist (although it could be, for the token-identity conception of statehood as well as the causal analysis of mental concepts do not rule out an immaterialist option). But it does seem to provide the right kind of internal material for it to be a two-factor approach to the intentionality of thought once it gets accompanied with the doctrine of externalism as it is commonly understood.

5. The representational theory of mind and cognitive psychology

There are several ways of saying what the representational theory of mind amounts to. For example, according to Nicholas Shea,

> it is the representational realism that has been deployed since the "cognitive revolution" in experimental psychology, cognitive neuroscience, and the other sciences of brain and behaviour. The central insight derives from the invention of mechanical computers, which gave us the idea that mental representations are physical particulars that are realized in the brains (and maybe bodies) of thinkers and interact causally in virtue of non-semantic properties (e.g. "form"), in ways that are faithful to their semantic properties. Psychological processes like thinking, perceiving, reasoning and imagining then consist of causal processes taking place between representations with appropriate contents. (Shea 2013: 498)

David Pitt expresses the main idea of representationalism in the following way:

> RTM *defines* [...] intentional mental states as relations to mental representations, and explains the intentionality of the former in terms of

the semantic properties of the latter. (Italics – M.G.) (Pitt 2012: §1)

Meanwhile, David Pickles links the theory to functionalism:

> The Representational Theory of Mind (RTM) holds (roughly speaking) that to have an intentional state, a state with content, is to be in a functional relation to an internal representation or 'vehicle' of that content. (Pickles 1994: 252)

Eric Margolis and Stephen Laurence note, that "RTM is standardly presented as taking beliefs and other propositional attitudes to be *relations* between agents and mental representations". However, they add,

> given that the relation in question is a matter of an agent's having a token representation with a particular type of functional role, it is simpler to say that beliefs just *are* mental representations with a characteristic type of functional role. (Margolis & Laurence 2007: 588, *f.*2)

First of all, it is important to bear in mind what the above passages clearly indicate: namely, that the representational theory of mind is *not* a theory in sub-personal cognitive psychology. It does not define or talk about sub-personal states. Rather, they all employ concepts like *thinking*, *perceiving*, *reasoning*, and *imagining*, having *an intentional state, a state with content*, or that of agent's *beliefs*. Accordingly, it should suggest that it is not sufficient that a theory subscribes to the idea that mental representations are empirically necessary for sub-personal psychology to explain cognitive capacities in order for a theory to count as representational theory *of mind*. For one to defend the representational theory of mind, it is also necessary to take an *additional* step and subscribe to the idea that mental states (thinking, reasoning, etc.) *are* representational states that are postulated in *sub-personal* cognitive psychology.

Second, the above descriptions of RTM should be distinguished from a different idea that is sometimes expressed by the same name. For example, Fodor maintains that

the point about propositional attitudes is that they are *representational* states: Whatever else a belief is, it is a kind of thing of which semantic evaluation is appropriate. Indeed, the very individuation of beliefs proceeds via (oblique) reference to the state of affairs that determine their semantic value; the belief that it is raining is essentially the belief whose truth or falsity depends on whether it is raining. (Fodor 1984: 232)

Moreover, he adds, "the notion of representation is crucial for every friend of propositional attitudes, not just the ones (like, say, Field, Harman and Fodor) whose views commit them to quantification over symbols in a mental language" (ibid.), so "realists about propositional attitudes are *ipso facto* Realists about representational states" (ibid.).

However, if the only reasons why propositional attitudes are representational states are that they are the "kind of thing of which semantic evaluation is appropriate" or that the very individuation of beliefs involves reference to their content, then the notion of representation has a different semantic force. For if "to represent" means nothing more and nothing else than, roughly, to have content, then it is important to bear in mind that the notion thus used should be disassociated from the idea of a *medium* – the idea that is standardly tied to the notion of representation. For it is not obvious that if *x* has content then *x* is a representation-qua-medium or that it involves one. For example, pictures (at least some of them) *have* content, and *are* representations, but it is not obvious that, say, thinking that *p* which has content must *be* a representation in the same sense in which a picture is (if that makes sense) or involve them.

Furthermore, the minimalist notion of representation, according to which being a realist about propositional attitudes implies being a representationalist, does *not* give support to the view that representations *in this sense* are mental representations that are being postulated by sub-personal cognitive psychology. In fact, if the representational theory of mind is to be based on a minimalist notion of "representation", then it would be not a controversial philosophical

theory about the nature of mind, but an uncontroversial claim which follows from uncontroversial assumptions about the individuation of some group of mental states. In other words, as Fodor rightly notes, every realist about propositional attitudes indeed would be a representationalist *in this sense*. But what also needs to be made explicit is that no "cognitive revolution" in cognitive psychology or controversial assumptions in the philosophy of mind are necessary to demonstrate this.

However, the representational theory of mind as it is advocated in philosophical psychology and the philosophy of mind is not meant to defend ideas that follow from uncontroversial assumptions about the individuation of propositional attitudes. Rather, within the domain of philosophical psychology, representationalism is understood as a kind of theory which is supposed to provide a general account of the nature of mind. As noted by Kathleen L. Slaney and Timothy P. Racine,

> the representational theory of mind (RTM) view describes both a general conceptualization of the nature of thought as well as a set of theories of mind, each of which attempts to explain the nature of mental states and processes. Broadly, RTM postulates the existence of mental intermediaries between objects observed and the observing agents, these intermediaries being some sort of *symbolic representation of that which is observed*. (Italics – M.G.) (Slaney & Racine 2011: 78)

As it is rightly pointed out by Glock, the "cognitive revolution" was in part stimulated by "Chomsky's (1959) demolition of Skinner's behaviourist linguistics, which paid homage to Descartes' stressing the non-mechanistic, 'creative' aspects of human language" (Glock *forthcoming*: Ch. 3) as well as by "the phenomenal progress of computer science, which seemed to open the prospect of a scientific and ultimately mechanistic construction and explanation of cognitive processes" (ibid.). Early cognitive science was based primarily on the computer model of the mind and cognition. But the situation has changed since the 1980s, when studies of brain imaging gave rise to connectionism, which no longer conceived of the mind as the software of a

computer. Instead, connectionism proposed to model it as a kind of neural network to be found in the brain. One way or the other, contemporary cognitive science continues to conceive of cognition in information-processing terms, and for present purposes the relevant implication of this idea is the representationalist theory of *concepts*.

Margolis and Laurence explain the connection between RTM and representationalism about concepts in the following way:

> The view that concepts are mental representations takes as its starting point a version of the Representational Theory of the Mind (RTM). According to RTM, thinking occurs in an internal system of representation; occurrent propositional attitudes are token mental representations (i.e., mental particulars with semantic properties). (Margolis & Laurence 2007: 562)

Thus, "when someone arrives at the belief that her house needs a new coat of paint, RTM says that she comes to form a mental representation, one that represents her house and its state of disrepair" (ibid.).

The idea that concepts are mental representations is not new. According to Slaney and Racine, "although its roots can be traced to antiquity, the first formal characterizations were offered up by classical empiricists" (Slaney & Racine 2011: 78). Assuming that "classical empiricism" stands for what some call "British empiricism", this is also Fodor's view:

> The philosophy of mind assumed in traditional British Empiricism was Realist about the attitudes and accepted a form of RTM. (Very roughly, the attitudes were construed as relations to mental images, the latter being endowed with semantic properties in virtue of what they resembled, and with causal properties in virtue of their associations. Mental states were productive because complex images can be constructed out of simple ones.). (Fodor 1985: 91)

Contemporary advocates of the representational theory of mind retain a more general idea that propositional attitudes are relations to mental representations while rejecting the claim that representations are *images*.

Slaney and Racine provide a more detailed description which reflects "the hegemony of the cognitivist Representational Theory of Mind tradition in both psychology and philosophy" (Slaney & Racine 2001: 81) and that is worth quoting in full. In their view, representationalism about concepts is committed to the following claims:

> (1) concepts are *mental particulars*, (2) concepts are *representations* of substances, events, and relations that occur (or potentially occur) in the external (i.e., nonmental) world; (3) *a fortiori*, concepts are *mental representations*; (4) concepts play a *causal-mediational* role between a substance, event, or relation and some cognitive ability (e.g. categorization, inference, or thought, more generally) of the bearer of the concept; (5) concepts *contain knowledge* or *information* about the substances, events, and relations they represent; (6) concepts are tied in important ways to the particular cognitive abilities of *categorization* and *induction*; and (7) concepts share important relations with *language*, but are not, strictly speaking, dependent on language. (Ibid.)

It should now be sufficient to give a more detailed explication of the view that "RTM defines [...] intentional mental states as relations to mental representations, and explains the intentionality of the former in terms of the semantic properties of the latter" (Pitt 2012: §1).

First, the idea that intentional states *are* relations to mental representations amounts to two claims. The first is that according to the theory, intentional states are representational states that are posited in sub-personal cognitive psychology after the "cognitive revolution", and these states are considered to be necessary to explain cognitive capacities. The second is that intentional states' being conceptual – i.e. intentional states, being in which requires having concepts – amounts to their containing concepts-qua-mental representations as described in propositions (1)-(7) by Slaney and Racine. Consequently, the intentionality of mental states *thus characterized* is explained by the semantic properties of concepts *thus understood*. This, in turn, has direct implications for a theory of intentionality of thought. For as Fodor

notes, "if RTM is true, the problem of the intentionality of the mental is largely
– perhaps exhaustively – the problem of the semanticity of mental
representations. […] It may be that what one descries, just there on the farthest
horizon, is a glimpse of a causal/teleological theory of meaning" (Fodor 1985:
99).

In order to illustrate the way in which "the semanticity of mental
representations" feature in the overall account of intentionality of thought, one
could make use of Fodor's schemata. In his theory, a constitutive account of
the nature of propositional attitudes amounts to the following claim:

> *Claim 1* (the nature of propositional attitudes):
> For any organism O, and any attitude A toward the proposition P, there
> is a ('computational'/'functional') relation R and a mental
> representation MP such that MP means that P, and O has A iff O bears
> R to MP … (Fodor 1987: 16-17)

When given a more perspicuous representation, "Claim 1" would amount to
the following constitutive claim:

> "O As that P" is true if and only if (1) and (2) are true.
> (1) O stands in a "computational"/"functional" relation R to a MP.
> (2) MP means that P.

Of course, the above claim is incomplete, for it does not provide an
explanation as to what determines the content "that P". In other words, it
assumes that MP *means* that P, but does not explain what determines that MP
means that P. For the orthodox semantic naturalist, clause (2) should be given a
reductive explanation in terms of factors that are non-intentional and are part of
some natural science, and it is the problem of providing a reductive and
naturalistic account of (2) that "the problem of the semanticity of mental
representations" amounts to.

It would be a mistake to conclude that a causal/teleological theory of
meaning requires RTM. It does not, and Davidson's anti-representationalist
causal theory of meaning – although different from the orthodox accounts in

not allowing for meaning and understanding to depend "on causal relations of which speakers may well be ignorant" – is a case in point (Davidson 2001: 151, 197-202). However, it does seem to be plausible to assume that *if* RTM is assumed, then, as Fodor proposes, the problem of intentionality of mental states should take the form of some causal/teleological theory of content.

As Speaks point outs, RTM implies that "any agent capable of having beliefs must have mental representations which are related in a certain way to the environment of the agent" (Speaks 2006: 434). More particularly, in his view, it implies that

> a mental representation has a property as its content just in case that representation bears *R*, a certain kind of causal relation, to the property. Such a theorist is then committed […] to the claim that any possible believer must process information in this way: by having certain parts of her cognitive system be *R*-related to parts of her environment. (Ibid.)

But one should take care. The claim that any possible believer *must process* information by having certain parts of her cognitive system be 'R-related' to parts of her environment not only indicates that it is a *constitutive* claim as to what *must* be true of *any* believer in order for it to have intentional states. It is also important to take notice of what gets concealed in this formulation: namely, that when certain parts of an agent's cognitive system are activated, and the agent is "processing information", this involves a series of "tokenings" of mental representations that *are* concepts.

If concepts *are* defined on the lines explained above, then one should make explicit what implications this view has for the nature of mental states. The main commitment that seems to follow from this account of concepts has been pointed out by John McDowell. According to McDowell, if one thinks of concepts as mental representations "in the sense in which, say, drawings or sentences are representations" (McDowell 1998a: 285), then being in a mental state that is representational *in the minimalist sense* "must in itself consist in the presence in the mind of an item with an intrinsic nature characterizable independently of considering what it represents" (ibid.). According to

McDowell, the implication follows because if concepts are mental representations in the sense in which drawings or sentences are, and "a representation is an item whose intrinsic nature is characterizable independently of its representational properties: a symbol" (ibid.), then one's being in mental state having of which requires concepts *thus conceived* implies that it contains items whose nature can be described independently of its representational properties. And if they *can* be thus characterized, then as for any representation – mental or not – one must provide an explanation as to how *they* get to represent the world. Putnam, who relied on the representationalist view of concepts in *Reason, Truth, and History*, has maintained that one cannot assume that thoughts are "intrinsically referential" without falling into what he called a "magical theory of reference" (Putnam 1981: 3-21). Naturally so, because if concepts *are* mental representations, one cannot assume that they have intrinsic representational (referential) properties without falling into a magical conception of representation.

Representational properties of language can be explained non-magically by its use, but what Fodor calls "the semanticity of mental representations" – the fact that concepts have semantic properties – cannot be explained in this way, for concepts' being mental particulars in the head cannot be used. The semanticity cannot be explained by means of resemblance for the reasons already noted (see Chapter 1). So given the representationalist conception of concepts, there really does seem to be no other way except to commit oneself to some form of causal theory of content, broadly conceived.

But it also should be clear that a commitment to RTM brings in a yet another internal element that is supposed to be present in thought, and that needs to be related to the things outside the head for intentionality to be possible *on this conception of concepts and mind*. And yet again: it might not be Cartesian enough – not pure enough – for some tastes as in its current mode RTM is not immaterialist. But irrespective of its commitment to some form of materialism, it does seem to have the resources to provide a yet another entity that is wholly in the head, and that is of the right kind for a theory that is

committed to RTM to amount to a bifurcationist picture of intentional attitudes when the external elements that are necessary to relate what is wholly in the head to what is outside the head are supplemented.

III. Concepts, agency, and the nature of mind: critical remarks

1. Concepts and the mind

1. 1. Representationalism about concepts

In the contemporary philosophical literature on concepts, it is common to distinguish between three ontological views about their nature. The orthodox view is the representationalist theory of concepts, the core idea of which is that concepts are mental representations in the heads of individuals. The second view is a Fregean theory of concepts, according to which they are Fregean *Sinne* that are abstract objects, and thus not part of the spatio-temporal world. Consequently, they are neither inside, nor outside the heads of individuals. According to the last view, which is usually called the "ability approach", ontologically speaking, concepts are neither abstract objects, nor mental representations, but abilities of a special kind.

However, even if the above classification is accurate with respect to ontological matters, it does not cut the boundaries in the same way in other respects. For example, neither representationalism, nor the ability approach rules out the view that concepts should be *individuated* as finely as Fregean *Sinne*. Hence, both views can be made compatible with *this* Fregean idea even if neither of them is compatible with Frege's view that concepts are abstract objects residing in a Platonic "Third Realm". More generally, when it comes to individuation conditions of concepts, all three *ontological* views are compatible with the Fregean criterion for the individuation of the contents of thoughts. For example, neither John McDowell, nor Gareth Evans or Anthony Kenny accept the Fregean ontology of concepts, yet all of them agree with

Frege on the need to distinguish between *Sinn* and *Bedeutung*, as well as to the criteria for the individuation conditions for content.

There are several reasons to reject the Fregean view of the ontology of concepts. One of them might be a rejection of the objectivist idea, according to which concepts exist independently of human minds, as self-subsistent abstract entities. Another one might stem from a commitment to extensionalism or nominalism, which rejects all abstract entities. For extensionalist, any abstract objects (concepts, properties, sets) should be part of one's ontology only if extensional criteria of identity for them are possible.

No scientific semantic naturalist believes in abstract objects, for if they did, the whole agenda for naturalization would hardly make sense. So it should be safe to conclude that no semantic naturalist is committed to the Fregean ontology of concepts. However, given the possible options, the orthodox position is the representationalist view. But it is would be a mistake to suppose that the adoption of this conception is merely a result of a commitment to the scientific form of naturalism, even if that would imply that the theories of semantic naturalism are to be in close relation to the one's proposed by some natural science. For the representational theory of mind, as well as a representational theory of concepts, as a philosophical theory about the nature of mind and concepts, are commonly defended on explanatory grounds. Two arguments that are commonly given to support it will be considered.

One argument for representationalism appeals to the alleged ability of the theory to explain the so-called productivity of thought, where the notion "refers to the fact that, under suitable idealization, there is no upper bound to the range of semantically distinct thoughts" (Margolis & Laurence 2007: 563). According to Margolis and Laurence,

> RTM's explanation of productivity is in terms of the compositional rules that govern the domain of mental representations. A finite stock of basic concepts can be combined over and over again, using the same rules to create increasingly complex mental representations. (Ibid.: 563)

So according to the representationalist position, "without mental representations, and without their having compositional structure, a fundamental fact about our minds would be left unexplained" (ibid.), a "fundamental fact" meaning a cognitive ability to "produce thoughts" in the sense explained above.

A second reason for representationalism is that it is alleged to be able to explain "how mental processes can be rational and yet realized in a physical system" (Margolis & Laurence 2007: 564). The answer, in this case, is "given in terms of the dual life that structured mental representations possess. They have physical-causal (presumably neurological) properties but they also have semantic properties" (ibid.). Hence, the idea that some mental representation can be described as possessing properties of two different kinds can explain how mental processes can be *both* rational and "realized in a physical system". The assumption is that without mental representations the fact that our cognitive processes are both rational and realized in a physical system would be left unexplained.

At this point, there are two reasons for being skeptical about the representationalist line of thought.

For one, if what is meant by the idea of "productivity of thought" is that one can (is able to) think an unlimited number (in some sense) of new thoughts, then it is not clear why postulating mental representations is necessary. For it is not clear why it is not sufficient to explain the ability to think more complex and new thoughts by reference to the development of simpler abilities to more complex ones. Unless, of course, one assumes that to have such an ability one must have some mental representation. But in that case, there is no need to argue for the postulation of mental representations by reference to the "productivity of thought", although of course a different argument should be provided as to why mental representations are necessary for having an ability.

Second, what is presented as an argument for a philosophical theory of the mind is in fact an argument for an empirical theory of cognition in

cognitive psychology: more specifically, a theory (or *part* of a theory) whose primary aim is to provide a causal-enabling explanation of certain cognitive facts and *not* a constitutive theory of the nature of cognition and concepts. Of course, there *is* a sense in which both kinds of theories concern the same subject matter: namely, both of them aim to provide an explanation – yet, as should be clear by now, of a *distinct* kind – of the facts about cognition (among other psychological phenomena). Moreover, the idea that explanations are of distinct kinds does not entail that they are independent and can be pursued in isolation. But if they *are* of distinct kinds, it rules out the idea that empirical theories and explanations in empirical psychology are on a par with constitutive theories and explanations in philosophy of mind.

There is yet another reason why arguments for representationalism in cognitive psychology do not *directly* support the representational theory as a philosophical theory about the mind, and it lies in a widely dismissed categorical distinction between personal and sub-personal states or facts.

One way to formulate the general theoretical demand of cognitive psychology is to maintain that it aims to explain how facts about the mental life of agents – broadly understood – are possible. Given that cognitive psychology is an empirical science, the notion of *possible* concerns *empirical possibility*: that is, empirical discoveries and theory construction that concern causal-enabling conditions of the relevant psychological facts. Some of the particular task of cognitive psychology include furnishing empirically adequate explanations of how agents can learn a natural language, or are able to form rational judgements, hold beliefs, perceive objects in the environment, perform intentional or non-intentional movements, or other complex cognitive or behavioral tasks. So as indicated earlier, there *is* a sense in which the subject matter of cognitive psychology concerns the mental life of agents, and so their explanations are directed at the relevant *personal-level* mental phenomena like reasoning, thinking, perceiving, etc. However, the *explanans* of cognitive psychology – i.e. the domain by reference to which the relevant facts are to be explained – concerns the *sub-personal* level. Thus, in order to explain how a

person can have the ability to perform certain cognitive or behavioral tasks, one refers to sub-personal facts of a person. But the distinction between the personal and the sub-personal level must be explained in more detail.

Daniel Dennett is considered to be the first to draw the distinction between the personal/sub-personal levels in his *Content and Consciousness* (1969). As he notes, "in one respect the distinction between personal and sub-personal levels of explanation is not at all new" (Dennett 1969: 95) as it is already implicit in the works of Wittgenstein and Ryle. The distinction is meant to differentiate between "the explanatory level of people and their sensations and activities" and "the sub-personal level of brains and events in the nervous system" (ibid.: 93). As it is pointed out by Jennifer Hornsby, "by making a distinction between levels of explanation […] one ensures that the kind of explanation distinctive of people and their sensations and activities is not confused with explanation of a different kind" (Hornsby 2000: 7).

If it is agreed, that so-called folk psychology concerns personal-level facts, whereas empirical psychology is primarily directed at the sub-personal level (at least with respect to its empirical-cum-explanatory theories), one then has the conceptual means to argue that any kind of argument for a theoretical posit in sub-personal empirical psychology *does not* imply that they are what the *personal*-level concepts or categories apply to. Accordingly, arguments against theoretical posits in empirical psychology do not imply eliminativism with regards to the phenomena at the personal level. Thus, Fodor's animadversions notwithstanding, rejection of representationalism in cognitive science *does not* imply eliminativism in philosophy of mind *if* the personal/sub-personal distinction is kept in mind.

However, the distinction is rarely kept in mind, both in contemporary philosophy of mind as well as in cognitive psychology. For example, Hornsby notes that

> when students of subpersonal Psychology speak of states possessing kinds of content which obviously do not feature in our thought, we may not be much inclined to think that *our* states have these contents: it

could not even sensibly be said that we (for instance) believed them. But when the Psychologists speak of states whose contents are of the right sort to be contents of states of ourselves, it is possible to forget that their concern is with the subpersonal level. (Hornsby 1997: 165-166)

But it is not only psychologists who can forget their concerns. Martin Davies proves to be a good example of how easy it is to forget (or ignore) the distinction, and switch between the two levels in the philosophical domain. He writes:

We can begin from the assumption that personal-level events of conscious thought are underpinned by occurrences of physical configurations belonging to types that figure in the science of information-processing psychology. These physical configurations can be assigned the contents of the thoughts that they underpin. So we assume that, if a person consciously or occurrently thinks that p, then there is a state that has the representational content that p and is of a type that can figure in subpersonal-level psychological structures and processes. (Davies 2005: 370)

By starting from the assumption that personal-level events are underpinned by sub-personal occurrences which, as he rightly notes, belong and "figure in the science of information-processing psychology", Davies then claims that the *sub-personal structures* can be assigned the content of thoughts that they underpin. So if *x* is a personal level "event", and *y* is a sub-personal "occurrence" that underpins it, then, according to Davies, the fact that *x* has content *p* allows to maintain that *p* can be ascribed to *y*. But this follows only if *x is y*, which rests on the assumption that there is no distinction between the personal and the sub-personal level.

But even if the best empirical theory of cognition must be based on *some* version of representationalism, there are at least two independent arguments against the idea that *concepts* are mental representations.

The first concerns the shareability of concepts. Recall that according to representationalism, concepts are a "kind of mental particulars" that are in the heads of individuals. However, if concepts are mental particulars, then this

implies that they cannot be shared. This implication would not be a problem if representationalists were willing to deny that shareability is a necessary condition for a theory of concepts. But that is not the case. For example, Fodor maintains that shareability is a "non-negotiable condition on a theory of concepts" and "concepts are public; they're the sort of things that lots of people can, and do, share" (Fodor 1998: 28).

The idea that concepts must be shareable was the main reason why Frege distinguished between senses (*Sinn*) and ideas (*Vostellungen*), and held that shareability of concepts can be explained by construing them as abstract – as opposed to concrete – objects. But Fodor does not seem to agree with Frege on this point, and holds that RTM can account for the shareability of concepts by introducing the type/token distinction. He writes:

> I'm assuming, in the general spirit of Representational Theories of Mind (RTMs), that the mental particular that's in your head on occasions when you think dog is a token of the concept type DOG, just as the word that's on your lips when you say "dog" is a token of the word type "dog". In both cases, the tokens are concrete particulars and the types are abstracta. Likewise, the mental particular that's in your head when you think that (judge that) dogs bark is a token of the mental representation type DOGS BARK. (Fodor 2003: 13, *f*.7)

It is true that Fodor's proposal can account for the shareability of concepts, but that comes at a price. As pointed out by Glock, "the type/token distinction cannot be used to invalidate Frege's argument, since it implies abandoning the claim that concepts themselves are particulars" (Glock 2010: 311). Assuming the type/token distinction, the shareability of concepts can be explained only by reference to concept-*types*, and not their *tokens*. But in that case the proposal that concepts are particulars would apply only to the *tokens* of the relevant types. Thus, if concepts *themselves* are to be identified with types, as they should be, then shareability of concepts can be explained, but at a price of giving up the representationalist idea that concepts *themselves* are particulars.

The second argument concerns the distinction between concepts as *aspects* of content (*what* is V*ed*) vs *means* of representation, and can be illustrated by reference to McDowell's comment on Burge's discussion of *de re* beliefs.

According to Burge, only *de dicto* beliefs are semantically speaking fully propositional and epistemically speaking "fully conceptualized". The reason for this is that, in his view, only *de dicto* beliefs are composed of semantic elements that are present in believer's "conceptual repertoire". Thus, Burge maintains, "traditionally speaking, concepts are a person's *means* of representing objects in thought" (italics – M.G.) (Burge 1977: 345). But as McDowell notes, Burge makes an illegitimate move when he shifts "from concepts are parts or aspects of the *content* of a representational state, such as a belief, to concepts as *means* of representation" (McDowell 1998a: 218). The *means* by which one thinks what one thinks are not *what* one thinks, as neither are the *means* by which one speaks part of *what* one says. So the argument runs as follows:

(1) If concepts are representations, then they are means of representation.

(2) If concepts are aspects of content, then they are not means of representation.

(3) Concepts are aspects of content.

(4) Therefore, concepts are not representations.

So there are some good arguments against the view that concepts are representations, mental or otherwise; some reason to be skeptical of the claim that they must be posited; and an argument, based on the distinction between sub-personal/personal levels, against the view that the entities, of whatever sort, postulated in sub-personal cognitive psychology have any implications for the subject matter of personal level psychology or the philosophy of mind.

Given the three alternatives with regards to the ontology of concepts, and assuming that neither the Fregean ontology of concepts, nor the representationalist alternative are plausible, it should be natural to examine the plausibility of the remaining alternative – the ability approach.

1. 2. The mind as a set of powers

The alternative view of the nature of concepts is based on a broader conception of the mind that is considered to be an alternative to the predominant representationalism, which is Cartesian in spirit, if not in letter. For example, J. S. Mingers notes that the predominant paradigm in information systems and artificial intelligence

> is Cartesian representationalism based on a split between mind and body, and a model of cognition as the processing of representational information. […] Cartesian representationalism formed the backbone of the main cognitivist period which started with meeting between Simon, Chomsky, Minsky and McCarthy in 1956 and continues, to this day, to be the "normal" paradigm within IS and AI. (Mingers 2001: 104)

Contemporary forms of "Cartesian representationalism" need not be based on the "split between mind and body" (or rather between mind and brain), and the reason why contemporary forms of Cartesianism need not imply dualism is that the idea of Cartesian approach has undergone a change: while abandoning the dualistic ontology of Descartes, it nevertheless retained some of the central, if implicit, aspects of the abandoned framework. As Glock notes, "in the wake of Descartes, the mainstream Western philosophy has treated 'mind', its equivalents and cognates as the label of a special kind of thing, whether it be a separate mental substance, as in dualism, or the brain, as in materialist monism" (Glock *forthcoming*: Ch. 3).

If Glock is right, then the idea that follows from contemporary Cartesianism is not that the mind is non-physical or non-material, or that the mind is *distinct* from the body or from the brain. Rather, what follows is the idea that the mind is a kind of thing or object: an idea that is compatible with both dualistic and monistic (idealistic or materialistic) ontologies.

The alternative, which rejects both representationalism and Cartesianism, is based on the Aristotelian framework, which gives a central role to the categories of *ability*, *capacity* and *power*. According to the capacity approach,

"'Mind' is no more the name of a *thing* than 'space', 'time', 'habit' or 'influence'" (White 1972: 464-465). Instead, according to this view, it should be regarded as a kind of *potentiality* or *power*. Powers are genuine attributes that can be possessed by particulars or substances. However, at the same time "a potentiality should not be reified, treated as a thing of a peculiar kind that somehow co-exists with the particular or substance that possesses it" (Glock *forthcoming*: Ch. 3). Therefore, "a power is neither a flimsy actuality nor an ethereal substance" (ibid.).

Furthermore, as noted by Kenny, "abilities and capacities are individuated by their possessors and their exercises, but they are distinct from both" (Kenny 2010: 106): capacities are not identical to what possesses them, nor should they be identified with their exercises. Lastly, they should also be distinguished from their vehicle: that is, from physical ingredients, structures or actualities in virtue of which the possessor – a particular or a substance – has the power. As Kenny notes, "a vehicle is something concrete, something that can be located and measured. An ability, on the other hand, has neither length nor breadth nor location" (ibid.: 108). Moreover, "the difference between a power on the one hand, and its exercise or vehicle on the other is a *category* difference, not a difference like that between solid and shadow" (ibid.).

One might wonder how the ability view of the mind differs from the doctrines of behaviorism, mind-brain identity theory or functionalism. This can be partly explained by reference to the distinctions between the *exercise* of a power, *conditions* (both opportunity as well as enabling ones) that must obtain for the power to be exercised, and their *vehicles*.

If the appeal to capacities or powers is part of a constitutive account of the nature of mind, and if powers are to be distinguished from their exercise conditions, then behaviorism – at least of the reductive kind – is a form of *exercise* reductionism. If the mind-brain type-identity theory implies that "the property of believing *p* is constituted by a certain first-order non-intentional property of agents (e.g. the property of being in a certain brain state)" (Speaks

2003: 58), yet first-order properties are the vehicles of powers, then the type-identity theory is a form of *vehicle* reductionism. As Kenny notes, "the whole point of distinguishing between powers and their vehicles was to separate conceptual features of powers from the empirically necessary conditions of their exercise" (Kenny 2010: 107). Thus,

> in the case of the mind the connection between the capacity and its exercise is a conceptual one: one could not understand what the mind was if one did not understand what kinds of things constitute the exercise of mental capacity. The connection between capacity and vehicle, on the other hand, is a contingent one, discoverable by empirical science. (Kenny 1989: 74)

In this context, one could maintain that behaviorism is right about the logical or conceptual connection between the mind and behavior, whereas the identity theory is right about the domain which constitutes the vehicle of the capacity that is to be investigated by the relevant empirical sciences. At the same time, however, mentality should not be identified with either its behavioral manifestations, or its vehicle.

The relation between the ability view and the doctrine of functionalism is more complex. Recall that functionalism is a theory which implies that "the property of believing *p* is constituted by a second-order property of first-order states of agents (e.g., the property of being in some state with such and such relational properties)" (Speaks 2003: 58). But if so, then it might seem that the ability approach is no distinct from functionalism *thus understood*. However, there are at least two important differences between the two views.

The first is that according to functionalism, mental properties are properties of first-order *states* of agents, and not of agents themselves; thus, the property of V*ing* that *p* is a property of a brain state (although "*state* of the brain" or "brain *state*" seems to have very specific connotations in this context, as explained in Part II, Chapter 4, section 4.2.), *not* of an agent. The ability approach does not imply that mental properties are properties of their vehicles: cognitive and conative capacities are properties of agents, and not their parts.

So even if the vehicles of cognitive and conative capacities do have powers *of their own*, neither cognitive, nor conative powers, according to the ability view, are powers of "first-order states of agents".

The second difference is that according to the standard form of functionalism, mental properties are constituted by second-order properties of first-order properties that *ground the abilities* of agents[9]. For example, when defending the ability approach in epistemology, Hyman writes that according to functionalism, "*knowing a fact is having the second-order property of having some first-order property that grounds the ability to be guided by it*" (Hyman 2015: 184). According to Hyman's description, functionalism implies that mental properties are distinct from *both* first-order (non-intentional) properties *as well as* abilities that are grounded by the first-order properties. But if the implication is correct, then the ability approach seems to offer a simpler empirically adequate hypothesis as to what mental phenomena consists in. For if the mind is what the first-order properties *ground*, then appeal to second-order properties of first-order properties when explaining the nature of mental properties is otiose. This is not to deny that causal-functional roles of complex neural interconnections might be useful in describing and explaining the workings of the *vehicles* of mental properties. The crux of the matter is that according to the ability view, these causal-functional roles are *not* what is definitive of mental properties, and the occupiers of these roles are *not* what mental concepts apply to. If it is a rational principle that simpler empirically adequate hypothesis is to be preferred, then the ability approach is preferable to functionalism.

A possible objection to the ability view is based on its explanatory futility: one could maintain that the ability view is useless, for it does not explain *why* cognitive agents have the abilities that they in fact possess. In a

---

[9] Despite the fact that according to functionalism, mental properties are second-order properties of first-order states, it should be kept in mind that this does not preclude them from identifying mental states with neural states. As Lynne Rudder Baker notes, "although functionalists define the types of mental states in terms of their causal and functional roles, they usually take the particular thing that occupies a given role (a "realizer") to be a brain state" (Baker 1995: 10), and even if the type of mental state is not to be couched in physical terms, "most functionalists identify particular mental states with first-order "realizer states"" (ibid.: 10-11).

sense, a theory *can* explain why some cognitive agents have the abilities that they have. But whether an explanation is successful depends on what kind of explanation is required. For example, if knowledge is a kind of ability, then one can explain why *a* is able to Φ by saying that it is because *a* knows that *p*, or something close to this. But the 'because' would be analytic or conceptual, and therefore would not be able to provide a satisfactory answer the causal question as to why *a* has the ability in question. But the fact that it cannot does not imply that the theory is false, because no constitutive theory of X is meant to purport a causal explanation of X. Consequently, if the objection against the ability approach is to be interpreted as meaning that the theory does not provide a causal explanation as to why cognitive agents have the abilities would miss the point. For as Glock notes, "it should be born in mind that the capacity approach does not purport to furnish a causal explanation of the phylo- or ontogenesis of mental phenomena […] or of the proximal (or neurophysiological) mechanisms that constitute their vehicle" (Glock *forthcoming*: Ch. 3). Hence, the fact that the ability approach does not furnish a causal explanation would be a reason to reject the theory only if it was meant to furnish one. But neither the ability approach, nor constitutive theories more generally, should be aiming at this. Hence, "appeal to capacities or potentialities more generally would be a bad science, yet it may nonetheless be good philosophy" (ibid.).

1. 3. Concept possession as a normative ability

When applied to the ontology of concepts, the ability approach comes down to the idea that "a concept is not an entity [...] but a disposition or capacity" (Price 1953: 320, 348); more specifically, "capacities exercised in acts of judgement" (Geach 1957: 7). The line of thought that generally leads to the idea that concepts are a kind of dispositions or capacities usually begins from

considerations regarding concept possession. For example, Michael Dummett, one of the defenders of the ability view, maintains that

> philosophical questions about meaning are best interpreted as questions about understanding: a dictum about what the meaning of an expression consists in must be construed as a thesis about what it is to know its meaning. (Dummett 1993: 14)

A similar approach has been also defended by Christopher Peacocke. In his view, "there can be nothing more to the nature of a concept than is determined by a correct account of the capacity of a thinker to what it is to possess that concept" (Peacocke 1992: 5), and he characterizes this principle as "the concept-theoretic analogue of one of Dummett's principles about language" (ibid.). Hence, according to Peacocke, "as a theory of meaning should be a theory of understanding, so a theory of concepts should be a theory of concept possession" (ibid.).

The indirect approach to the nature of concepts explicates the fact that concepts have an intrinsic *epistemic* dimension. In the same way as it is part of the established notion of linguistic meaning that the meaning of an expression must be known to competent speakers who have mastered it, it too is a part of the concept of concept that they must be known by those, who possess and employ them. Hence,

> what concepts subjects possess or employ is manifest in their cognitive and linguistic operations and achievements, especially in how they employ and explain the corresponding terms. Therefore, we can establish what concepts a subject possesses or employs by looking at her cognitive and linguistic activities and at the way she justifies and explains them. (Glock 2010: 98)

Furthermore, as Kenny notes, "a sufficient, but not a necessary condition for a person to possess the concept of F is that she shall have mastered the use of a word for 'F' in some language" (Kenny 2010: 106). If Kenny is right, then concept possession is (or is closely linked to) linguistic understanding.

But if concept possession as well as linguistic understanding is a kind of capacity or ability, the question remains as to what kind of ability it is.

One possible answer is that concept possession is the capacity to recognize or discriminate between different types of objects (see Price 1953; Dupré 1996: 331). But as some philosophers have objected (Geach 1957: 16-17; Davidson 1997: 25), this position implies an intuitively implausible view that the discriminatory capacities of lower level animals are sufficient for concept possession. So concept possession should require more than *mere discrimination*.

There is disagreement over what more is needed. But those who accept an essentially normative character of conceptual thought tend to follow Davidson (1997: 24-25) in holding the view that concept possession requires the ability to *classify* things. Glock points out that "to be capable of classifying things, a creature A must not just have a disposition to respond differently to *F*s and non-*F*s. A must also be capable, for instance, of recognizing and correcting mistakes" (Glock 2010: 28).

If above considerations are correct, then "concept possession depends […] on discriminatory behaviour that is sufficiently complex and flexible to be guided by normative standards" (ibid.). In short, concept possession is a capacity to classify things as being a certain way, where that capacity involves the ability to recognize and correct mistakes, i.e. to be guided by *normative standards*.

Concept possession can also be given a description on a more general level: not as to *what* capacity it is, but as to what *kind* of capacity or ability it is. Most of the abilities of human beings are two-way powers, which contrary to one-way powers are *not* automatically exercised given antecedent conditions. For example, the ability to cycle is a two-way power because an agent can choose whether to exercise it or not: one can decide to take a train or walk even if antecedent conditions for an exercise of the ability to cycle are present. The ability to digest food, however, is a one-way power because it gets automatically exercised given the antecedent conditions. In short, two-way powers are distinguished from one-way powers by reference to the idea that the former are subject to the will whereas the latter are not.

However, it seems that concept possession is neither a one-way, nor a two-way power. First, it is not a two-way power for it is not subject to the will. As Kenny points out, "looking up at the flashing lights of the advertisements in Piccadilly Circus, one cannot prevent oneself from understanding their message" (Kenny 1989: 22; also 1975: 52-53). But concept possession is not a one-way power either. For concept possession is subject to *normative standards*, yet one-way powers cannot be subject to normative standards. The latter point has been emphasized many times in the literature on the nature of dispositions. On the assumption that dispositions are one-way powers[10], it is commonly maintained that dispositions lack normative force: a possessor that fails to act in accordance to a disposition cannot be subject to normative assessment and correction. Thus, concept possession "by contrast to two-way powers […] is not subject to the will, and by contrast to mere dispositions or one-way powers its exercise is subject to normative assessment: conceptual classifications are susceptible to criticism and correction" (Glock 2010: 322). Hence, if concept possession is to be construed as a kind of potentiality or power, then it sits somewhere in-between one-way and two-way powers.

One way or the other, the above considerations do not support the idea that concepts are abilities. So even if it is true that the categorical framework of concepts and meanings should be construed in terms of conditions of their possession, it does not follow that *what* is being possessed is an ability or capacity itself.

For example, Hacker maintains that "to possess the concept expressed by a word is to have mastered the technique of use that is common to all expressions, in the same or different languages, that have the same meaning – that have the relevantly equivalent use" (Hacker 2013: 129). So in Hacker's view, concept possession amounts to a technique of use of linguistic expressions. This, in turn, implies that concept possession is impossible unless

---

[10] At this point is should be noted that due to influences stemming from behaviourism, it is common to link the notion of a "disposition" to *one-way* powers and maintain that two-way powers are *not* dispositions. Accordingly, the question whether conceptual or linguistic capacities are dispositions depends, in part, on how the notion of a "disposition" is to be used.

one can master a natural language. But the idea that natural language is necessary for thought or concept possession is contestable, and the main arguments for the view that at least *some* concepts can be possessed without possession of natural language come from considerations with regards to animal cognition (see Kenny 1989: 36-37; also DeGrazia 1996: 154-156).

Be that as it may, at this point it is not necessary to take a position on whether the "lingualist" view is true, and it should suffice to maintain that concept possession involves a technique that at least with respect to *some* concepts requires mastery of natural language. But even if concept possession is not necessarily a technique of use of *linguistic* expressions, it still can retain the idea that it is a technique *of some kind*.

According to Glock, "in so far as conceptual thought involves a technique, it is a technique of operating according to a rule or principle" (Glock 2010: 326) which *may* yet *need not* be attributed to natural language expressions. But if conceptual thought is a technique to operate according to some rule or principle, it is natural to make a further step and conceive of concepts *as rules*. Thus, concept possession is a technique to operate according to a *rule* or *classificatory* principle, and in that respect concept possession is a *normative* kind of ability.

1. 4. Implications and theoretical virtues of the ability approach

One might wonder whether the ability view of concept possession, as well as a more general approach to the mind that is based on the categories of powers and capacities, is compatible with some of the predominant theories in foundational semantics. One of them is the doctrine of semantic externalism.

It should be clear by now that the doctrine of semantic externalism might have different faces, depending on what particular claims are being made when characterizing it. For example, if one considers a *general* characterization of the doctrine of externalism as provided by Bilgrami (1992: 2), then it does not

seem to be in any contradiction to the view that concept possession is a kind of capacity or ability. For it might well be true that one could not have any concepts, and no conceptual thought, if it turned out that there was no external world. It also seems to be reasonable to assume that intentionality cannot be wholly characterized independently of the external world, or merely by reference to what one *does*. But when one considers whether the ability approach is compatible with the standard version of externalism, according to which content is determined by external factors, most likely such that the agent is not aware of them, then it does seem to be incompatible with this idea.

The ability view maintains that having concepts and conceptual thought is having a capacity of a particular kind. Thus, even if in *some* cases relation to the external environment might be necessary for A to V that *p*, it is not sufficient. And no other external conditioning, however complex, should suffice. For according to the ability view, A's V*ing* that *p* implies that A should be able to *explain* what A V*s*: the content of A's V*ing* is as determinate as A's understanding of the constituent concepts, which requires that A would be able to explain them – to explain *what* A V*s*. This, in turn, seems to imply that at least with respect to *conceptual* thought, no causal conditioning is sufficient for A's intentional state to possess a particular content. If there is no appropriate response, if A cannot explain what A V*s*, then no matter what causal stimulations are irritating one's nerve-endings, there seems to be no fact of the matter as to what A V*s*, and no reason to hold that A V*s* anything at all. Of course, that should not be taken to imply that all V*ing* should be actually expressed, but what the ability approach does seem to require is that V*ing* should be *capable* of being expressed. Thus, the ability view seems to imply that the general answer to the content determination question is this: what a person thinks, believes or desires is determined by what a person would do or say in various circumstances. If reference to various circumstances is externalism enough, then the ability approach *is* an externalist one, although of a kind that differs from the standard versions defended by Putnam or Burge. But it is not compatible with the view that the contents of one's thoughts are

determined by factors external and unknown by the subject, or factors of which a subject is not aware. Of course, the ability view is compatible with the idea that one can be wrong about what a word means, as well as to the idea that one can misunderstand a word. But neither of these ideas seem to require that what one thinks or means is directly determined by "external" factors that are unknown.

Consequently, according to the ability view, the concepts of intentional states are not composite, and so a constitutive account of intentional concepts does not imply that "part of the complete truth about the mind is the truth about something wholly in the head; another part of the complete truth about the mind is the truth about how the subject matter of the first part is related to things outside the head" (McDowell 1998a: 278). On this approach, there is nothing that is constitutively relevant for mentality that is in the head, and so nothing needs to be said about how what is in the head relates to things outside the head in order for intentionality to be possible. Of course, the ability view must tell something about what it is for mental states to possess content. But if the mind is a set of abilities or powers, this question should not be understood as a question about there being *something* in the head and *its* relation to the world outside the head. In this respect, the ability view does not have a consequence of being committed to the bifurcationist picture of intentional states.

One should note at this point, however, that for Putnam commitment to externalism turned out to be a route to some form of the ability view: as he puts it, the mind is "a structured system of object-involving abilities" (Putnam 1992: 356). This, in turn, confirms the idea that externalism does not imply a bifurcationist account of the mind. In fact, it seems that Putnam adopted the ability approach to the mind in order to avoid the bifurcationist conception. In particular, he gave up on functionalism – a theory which, when combined with externalism, does seem to imply a bifurcationist picture of the mind. When commenting on the possible consequences of the "externalist turn", Richard Schantz writes:

externalist turn is what in the end motivated Putnam's rejection of his former functionalist theory of the mind. If propositional attitudes aren't inner states, the nature of which can be characterized independently of external factors, then the core thesis of functionalism, the thesis that mental states are algorithmic states in our heads, cannot be right. Functionalism ultimately fails because it embodies an internalist view of the mind. It fails, in other words, because the mind is not in the head either. (Schantz 2004: 17)

If Shantz is to be trusted, then it seems that the ability view might be compatible with externalism after all: Putnam has rejected functionalism, adopted the ability view and seems to remain an externalist.

One way to make sense of this possibility could be by reconceiving what externalism amounts to when the doctrine of functionalism that has guided the arguments for externalism has been given up. Perhaps externalism within the context of functionalism and "narrow" states amounts to something slightly different than it does once functionalism is rejected. In fact, it seems that this is exactly what happens with the core ideas of externalism once they are placed in a different picture of the mind. For example, McDowell, an externalist who is sympathetic to the ability approach to the mind, writes:

To begin with in "The Meaning of 'Meaning'", and in a number of writings since then, Hilary Putnam has argued trenchantly, and I think convincingly, that in the case of at least certain sort of words, *the environment of those who use them enters into determining their extension*. (Italics – M.G.) (McDowell 1998a: 276)

One might question the adequacy of McDowell's description of what has been shown by Putnam in a number of his writings, but even if his description is inaccurate, one might still consider McDowell's idea on its own. And if one takes it that way, and considers it to be a description of what externalism is, then it might be easier to conceive of how the ability view might be compatible with externalism *thus understood*. For the ability approach does not seem to imply that the environment cannot "enter" into determination of the *extension* of at least some terms or concepts, although the term "enters" should be spelled

out in a greater detail. But if it is spelled out in the way that does not conflict with the link between meaning, understanding, and being capable to explain or manifest a capacity in question, then perhaps externalism as characterized by McDowell *is* compatible with the ability approach.

Another pertinent question might be whether the ability approach is compatible with the ideas of informational semantics, biosemantics or other scientifically driven naturalistic theories of semantics. In order to answer this question, one would need examine specific ideas one-by-one. But on a more general level, the ability approach does not tell anything as to what might be empirically necessary for having certain abilities or capacities, or how certain functions work of have evolved, and the approach should be in line with developments in biology, cognitive ethology, empirical psychology, or other relevant empirical sciences.

This might not seem to be of importance to those who are in "horror of the normative", but it should be noted that the ability view *can* explain the normative aspect of semantics and propositional attitudes, which is a problem for causal theories of content, broadly understood. For if conceptual thought is a technique of operating according to a rule or principle – that is, if it is a normative ability – then it can explain the normative aspect of meaning or propositional content, which gets determined by an exercise of the ability of this kind.

Furthermore, the ability approach does not seem to face the self-knowledge problem, for it does not imply that content is determined by factors of which the subject is not aware. However, A's knowledge of what A V*s* is conclusive only in cases where A is not subject to linguistic or factual ignorance. As Glock notes, in cases where A is ignorant of some linguistic or empirical facts, "what people think does not just depend on what they avow to be thinking, or on what ascriptions they would accept, but also on how they would *explain* their thoughts when challenged, and on what knowledge they have of relevant facts" (Glock 2001: 115-116).

As for the fine-grainedness problem, it can allow for content to be as fine-grained as needed, and the "grain" should be as fine as A's explanations given of what A V*s*. The shareability of concepts too does not seem to be a problem. For according to the ability view, concepts are *abstractions* from techniques of use or classification. This does imply that concepts are abstract. However, the idea of their being abstract need *not* imply that entities which belong to this kind are denizens of the Platonic "Third Realm". For if the category of a concept is linked to the idea of concept possession, and if concept possession is a kind of ability, then concepts have a more subjective dimension than is allowed by objectivist forms of Platonism. At the same time, the idea of subjectivity involved does not imply that concepts are private entities. What it does imply is that concepts are tied to the cognitive abilities of a subject. As Glock notes, "both propositions and concepts are abstract and yet dependent on cognitive subjects capable of thinking and/or speaking. Both notions make sense only by reference to the activities and abilities of such subjects" (Glock 2010: 329). In short, "propositions and concepts are logical constructions from linguistic and cognitive abilities and activities" (ibid.: 324), and this view *is* compatible with the idea that different subjects can possess the same ability. Thus, the ability view does *not* have an implication that "it is improper to speak of two people as 'having the same concept'" (Geach 1957: 14).

The last, although perhaps to some the most crucial, question to be considered here has to do with the current modes of thought in philosophy: the question being whether the ability view is compatible with the doctrine of naturalism.

As it should be clear from the discussion of naturalism, the answer to this depends on what one means by "naturalism". Perhaps it would be better to consider what the ability approach implies, and then see whether that deserves the label "naturalism". For one, it does not imply that the mind is a super-natural entity, so it is a naturalistic position in this minimal sense. Furthermore, it seems to be naturalistic in the sense that it implies that the mind constitutes the explanandum of cognitive psychology. Recall Cain's characterization of

what happened after the "cognitive revolution". He writes that psychologists "came to see their central concern as being that of explaining intentionally characterized cognitive capacities and held that in order to explain such capacities, it is necessary to appeal to internal representational states and processes" (Cain 2002: 20-21). It is clear that the ability view is compatible with the first part of this thought; in fact, it implies that the central aim of cognitive psychology is to provide explanations of mental phenomena, which is exactly how it is generally conceived. However, the ability view need not be committed to the second part of the thought (after "necessary"), for it does not make any claims as to what is empirically necessary to explain such capacities. So the ability view is a naturalistic position in the sense of constituting the explanandum of cognitive psychology. However, it is not naturalistic in that cognitive capacities are neither to be understood in terms of their vehicles, nor in terms of concepts used on the level of sub-personal psychology.

On the other hand, there does not seem to be any good reason to hold that only what is at the sub-personal level, or at the level of vehicles *can* be called natural. Our being able to do things, having various kinds of capacities are as natural facts about this world as facts about our walking or talking. This does not mean that the story as to how it relates to everything we know about the world is going to be simple or that there is no story to be told. Quite the contrary: one does face the same kind of questions that are being posed for other theories about the nature of mind, although what those questions mean might differ, given a difference in a conceptual framework. But there does not seem to be any good reason to think that the solutions it provides are not "naturalistic" enough. The ability approach might not be the whole truth about the nature of mind, and it might be mistaken about one or another issue. But if some part of it, or even the whole framework is to be rejected, it should not be because it is not "naturalistic" enough.

1. 5. Descriptivism about Fregean *Sinn*

It should be clear by now that the ability approach to concepts and concept possession is rightly taken to be an alternative framework to the essentially Cartesian view of mentality and its relation to the world. But sometimes the abandonment of the Cartesian conception of the mind is taken to require a rejection of the Fregean principles of philosophical semantics. This issue is not directly related to the ability approach to the mind, although anyone attempting to develop an ability-based philosophical semantics on the basis of Fregean notion of *Sinn*, will face the problem of reconciling the non-Cartesian picture of mentality with the allegedly internalist notion of Fregean *Sinn*. The following sections aim to show that the internalist conception of Fregean semantics that is based on the notion of *Sinn* is misguided.

The main reason for the internalist interpretation of Fregean *Sinn* is based on descriptivist considerations about the nature of meaning and reference. As McDowell notes, "modern philosophical thinking about the relations of thought to objects was for a long time captivated by the extended Theory of Descriptions" (McDowell 1998a: 252-253) and so it is commonly taken for granted that "the notion of singular sense is half-baked forerunner of the Theory of Descriptions" (ibid.: 269). In McDowell's view, the main idea of the view can be expressed in the following way:

> It is commonly believed that a Fregean philosophy of language and thought can represent an utterance, or a propositional attitude, as being about an object *only by crediting it with a content that determines the object by specification*, or at least in such a way that *the content is available to be thought or expressed whether the object exists or not*. (Italics – M.G.) (Ibid.: 214)

So the descriptivist-cum-specificatory interpretation of Fregean semantics involves two ideas. First, it maintains that the *Sinn* of singular terms, which determines its *Bedeutung* by way of specification, is essentially descriptive. Second, the descriptive content is assumed to be insensitive to whether its

object – the *Bedeutung* – exists. It should be noted that a Fregean *Sinn* is not only supposed to be part of the meaning of linguistic expressions; in the Fregean semantics, it is also assumed to be what determines the *Bedeutung* of a term, yet *not* in the sense of content-determination question, but in the sense in which formal functions determine their values[11].

However, once Fregean *Sinn* is conceived on the descriptivist lines, then "a complete account of the mode in which an object is presented to us – the effect that it has on our cognitive representations or on our store of information – may be insufficient to determine that one object rather than another is the subject of our beliefs or statements" (Burge 1977: 357-358). And the main reason why *Sinn* is considered to be insufficient to determine its *Bedeutung* is because it is assumed to be equivalent to some definite description (or their cluster), where the latter are construed to be independent from what they are taken to refer to. Hence, the gap between *Sinn* and *Bedeutung*.

At this point it shall be noted that the descriptivist conception of reference or intentionality need not imply the idea that descriptive content needs to be *verbally* expressed in either definite descriptions or purely general terms. As McDowell notes,

> "Descriptivist" need not imply that the specification by conformity to which the object of a thought is supposed to be determined must be linguistically expressible. […] The point of the label "descriptivist" is to stress (by way of allusion to the Theory of Descriptions) the crucial point that these modes of presentation are not object-dependent. (McDowell 1998a: 257, *f*.57)

If McDowell is to be trusted, then the main idea of the descriptivist theory of meaning or reference is not that an expression *means the same as* some definite description or their cluster. The important point is that according to descriptivism *thus understood*, the expression has a kind of content that can be

---

[11] Of course, the fact *that Sinn* determines *Bedeutung* needs an explanation, and it should be part of a full theory of meaning or intentionality.

represented by means of the Russellian theory of descriptions, and which can be entertained even if the object that it is supposed to apply does not exist.

John Searle's theory of intentionality which he claims to be "Fregean in spirit, though, of course, not in detail" (Searle 1983: 197) is a good example of a descriptivist and essentially internalist account of the Fregean philosophy of language and thought. According to Searle, the *Sinn* of a singular term determines its *Bedeutung* by means of something like a specification or set of conditions that the right object must fit or satisfy. Moreover, "linguistic reference is a special case of Intentional reference" (ibid.: 197-198) while "Intentional reference is always by way of the relation of fitting or satisfaction" (ibid.: 198). In Searle's view, Fregean accounts of meaning and intentionality, including his theory, "are internalist in the sense that it is in virtue of some mental state in the head of a speaker and hearer – the mental state of grasping and abstract entity or simply having a certain Intentional Content – that speaker and hearer can understand linguistic references" (ibid.). But it is also internalist in a more general sense, in a sense in which it rejects a general characterization of externalism: in Searle's view, Fregean semantics implies that all beliefs or other mental states could be had by a brain in a vat (ibid.: 230).

Searle rightly notes that it is common ground among many philosophers working on philosophical semantics that the "exponents of the so-called causal theory of names and the causal theory of reference are supposed to have refuted something called the "descriptivist theory" of names and of reference" (Searle 1983: 199) which, in his view, entails that they "thereby refuted any internalist or Fregean account, and to have shown that reference is achieved in virtue of some *external* causal relations" (ibid.). But even if Searle is right about the causal theory of reference and its relation to the descriptivist theory of reference, it should be clear that if the descriptivist account of reference is assumed to be "internalist" in Searle's sense, it does not imply that a rejection of it is a rejection of Fregean semantics *unless* Fregean *Sinn* is understood on descriptivist-cum-internalist lines. On the other hand, rejection of the idea that reference is achieved in virtue of external causal relations does not commit one

to the internalist account of intentionality according to which reference is determined by a descriptive Fregean *Sinn*.

Be that as it may, there could be some good reasons to think of Fregean *Sinn* on the descriptivist lines. Here are some that have been taken to motivate the internalist picture of Fregean *Sinn*.

One reason for the descriptivist interpretation seems to be based on the idea that the basic motivation lying behind the semantic projects of Frege and Russell was the same. So Fregean *Sinn* is assumed to be a kind of category that would do the same explanatory job as definite descriptions do in Russell's theory of descriptions.

But the assumption is mistaken. The category of *Sinn* was introduced to secure a fineness of grain in object-oriented contents of language and thought that cannot be secured if content is individuated only by reference to objects it is about, as it is in Russell's (or a neo-Russellian) theory of propositions. The primary aim of the Russellian theory of descriptions, on the other hand, is to show how what seems to be a singular proposition can be significant in cases where the object that it is about does not exist. Frege did not explicate the distinction between *Sinn* and *Bedeutung* by reference to a theory of descriptions, nor did he maintain that *Sinn* is synonymous with some definite description. Furthermore, even if it is true that in some cases *Sinn* can be (yet need not be) indicated by a use of a description, this does not show that Sinn *is* descriptive or that it should be analyzed on the model of Russell's theory. The idea that definite descriptions are a poor model to understand Fregean *Sinn* has been defended several times (see Dummett 1973: 110-111; Bell 1979).

If the descriptivist interpretation of a Fregean *Sinn* is unjustified, then one of the main arguments against it is undermined too. For it is commonly held that Kripke's (1980) and Donnellan's (1970) attacks against the idea that having some definite description is either necessary or sufficient for a term to refer is also an attack on the Fregean notion of *Sinn* or Fregean semantics more generally. As argued by Kripke and Donnellan, descriptions are neither necessary, nor sufficient for reference. So if a Fregean *Sinn* is descriptive, then

Sinn is neither necessary, nor sufficient in determining the objects of thought or reference of linguistic expressions. But if the descriptivist account of *Sinn* is false, then Kripke's and Donnellan's arguments are not arguments against Fregean *Sinn*.

Another argument against a Fregean *Sinn* is based on the idea that *Sinn* implies an *indirect* view of reference which should be rejected. The alternative conception of reference that is *not* to be based on a Fregean notion of *Sinn* is usually known under the name of "direct reference". However, the term is unfortunate and it should not be taken literally. For proponents of the so-called direct reference view usually assume a causal theory of *reference determination* which does not seem to imply that reference relation is direct; in fact, quite the contrary seems to be the case on this view.

But the use of "direct" in this context has a particular connotation, and "being direct" is usually meant to imply here that reference is *not* mediated by a Fregean *Sinn*. This line of thought presupposes that the idea of Fregean *Sinn* implies indirectness. For example, according to David Kaplan, "the semantics of direct reference" simply means a set of "theories of meaning according to which certain singular terms refer directly without the mediation of a Fregean *Sinn* as meaning" (Kaplan 1989: 483).

But the idea of a Fregean *Sinn* does not imply indirectness, and having a Fregean *Sinn* is not a case of "mediation" between the term and its Bedeutung. As Gareth Evans notes, once

> we realize that the possession by a singular term of a Fregean sense can depend upon nothing more that its being associated with a proprietary way of thinking about an object, the idea that thought about an object which depends upon grasp of a Fregean sense must somehow be indirect will seem absurd. (Evans 1982: 62)

In other words, if the idea of a Fregean *Sinn* is, roughly speaking, an idea of a mode under which a *Bedeutung* is given, then it does not imply that *Bedeutung* is presented in some indirect way or that the relation between an expression and its *Bedeutung* is "mediated".

It is likely, however, that one of the reasons giving support to the idea that *Sinn* implies indirectness is Frege's ontology of concepts, according to which *Sinn* is an abstract object residing in a Platonic "Third Realm". So it might be true that *within Frege's ontology*, the idea of Fregean *Sinn* does imply the indirect conception of reference and intentionality. But Frege's ontology of concepts is not necessary for a *Fregean* view. For example, if the ontology based on the ability view is adopted, then the idea of being abstract does not require the Platonic "Third Realm", and thus gives no support to the claim that a Fregean *Sinn* is some intermediary object between the expression (or thought) and its *Bedeutung*.

Yet another argument against Fregean *Sinn* is based on an alternative reading of the notion of "direct reference" which links it to the Kripkean idea of rigid designation. As Kaplan notes, "I intend to use 'directly referential' for an expression whose referent, once determined, is taken as fixed for all possible circumstances" (Kaplan 1989: 493).

The idea of rigid designation is not exactly like Kaplan's intended use of "direct reference", for a term is supposed to be rigid or not in virtue of the fact that it does or does not have the same reference in other worlds, not in virtue of the fact that once its reference is determined, it is taken to be the same ('is fixed') across other possible worlds. In any case, the idea that *Bedeutung* is determined by *Sinn* does not preclude either the idea that *once determined*, the *Bedeutung* of a term can be held constant across other possible worlds, or the idea that when determining the truth-value of a sentence in modal contexts, the relevant component of a term is its *Bedeutung*, and not its *Sinn*. If so, then Fregean *Sinn* is compatible with either of the two interpretations of the notion of "direct reference" as well as the Kripkean idea of rigid designation.

If the above considerations are correct, then a descriptivist interpretation of Fregean *Sinn* is unfounded and criticisms of it that are based on the explicated ideas of direct reference or rigid designation are unjustified. Surely, that does *not* establish that a Fregean *Sinn* is *not* internalist – this should be argued on independent grounds. But as the above consideration indicate,

Fregean semantics need not be internalist or descriptivist, and need not imply an indirect view of reference or intentionality. Consequently, a theory developed on the principles of Fregean semantics needs not give support to the Cartesian or potentially bifurcationist accounts of meaning and mind.

2. Intentionality and action

2. 1. The problem of content epiphenomenalism

The relation between the problem of content epiphenomenalism and a theory of content is not direct: the latter is related to the former in virtue of presuppositions that transcend the field of the philosophy of language and mind, and belong to issues in action theory and mental causation. But the problem of content epiphenomenalism is of central importance to any theory of content – including semantic naturalism – that is externalist in character and which aims to be compatible with some of the common presuppositions regarding content's role in the explanation of action.

Here is how the problem is presented in *Dictionary of Cognitive Science*:

If, as the *physicalist* viewpoint holds, an individual's beliefs are the states of his/her brain […] Isn't the content of a belief rendered causally ineffective by the physical properties underlying the state of the brain? Don't these properties suffice for producing the bodily movement? Isn't the content of a belief epiphenomenal in the movement production process? (Houdé 2004: 50)

The problem of content epiphenomenalism is not restricted only to the doctrine of semantic naturalism: it faces *any* theory of content which shares the assumptions that give rise to the problem. The premises that imply the problem of content epiphenomenalism form an interconnected web that covers several areas of philosophy. But there are two central ideas that are necessary for the preliminary formulation of the way "the physicalist viewpoint" is commonly taken to pose a problem.

The first is the functionalist idea that beliefs and other mental states are functional states of a person's brain, so content properties are thus properties of those states. The second idea is an implication of semantic externalism: namely, the idea that content properties are *not* intrinsic properties, and thus are *not* neural properties of functionally specified brain states. Hence, if neural (physical) properties of states of the brain are sufficient for "the movement production process", then, it is commonly thought, content properties are epiphenomenal for "producing the bodily movement". As Fred Dretske puts it,

> if one assumed that bodily movements to the refrigerator are caused (in an appropriate way) by a belief and desire, these movements need not and, if we accept the scientific story, are not explained by what the person believed and desired. Not unless one assumes that the electrical/chemical/mechanical explanations given by neurophysiologists are – and will always be – deficient or incomplete. (Dretske 2009b: 14)

Thus, he concludes, "if intentional psychology, explanations of behavior in terms of a person's beliefs and desires, is competing with neuroscience for an explanation of bodily movement, then I, for one, don't see much hope for intentional psychology" (ibid.: 18).

It should be clear that the reasoning rests on the premise that having content is not a physical (neural) property – it assumes, that is, that the doctrine of semantic externalism is true – and can be represented in the following way:

(1) Neurophysiological properties are causally sufficient for causal explanations of bodily movements.

(2) Content properties are not neurophysiological properties.

(3) Hence, content properties are causally irrelevant (epiphenomenal) for explanations of behaviour.

It is of crucial importance to note that the conclusion of the argument follows – and so the "physicalist viewpoint" poses the problem – only if "bodily movements" (or its cognate) in (1) is equivalent to "behaviour" (or its cognate) in (3). In other words, (3) follows only if behaviour *is* a bodily movement.

160

Consequently, the argument for the problem of content epiphenomenalism works only if it is supplemented by an additional premise:

(1*) Neurophysiological properties are causally sufficient for causal explanations of bodily movements.

(2*) Content properties are not neurophysiological properties.

(3*) Hence, content properties are causally irrelevant for explanations of bodily movements.

(4*) Behaviour is a bodily movement.

(5*) Hence, content properties are causally irrelevant for explanations of behaviour.

Thus, as Dretske puts it, "if the neurophysiological story is correct, then actions must be something other than bodily movements if reasons are to explain them" (Dretske 2009b: 18).

But they must be something else – and not only because neurophysiological explanations are assumed to be correct and sufficient. The reason why actions (or behaviour) better *not* be bodily movements is that on the assumption that they are, intentional or folk psychology ends up competing with neurophysiology: as Jaegwon Kim indicates, "distinct explanations (in particular, causal explanations) of a single explanandum seem to exclude one another, in spite of the fact that their explanatory premises are mutually consistent" (Kim 1990: 38). This, in turn, shows that the problem of content epiphenomenalism depends on a theory of action: in Dretske's words, "if philosophers cannot agree about what action is, they cannot hope to understand the power of reasons to explain it" (Dretske 2009b: 13). That is, they cannot understand how content properties can be explanatorily relevant for explanation of action.

So the problem of content epiphenomenalism is a consequence of three incompatible ideas:

(1**) Neuroscientific explanations (that are not deficient and are not incomplete) of bodily movements *do not* cite reasons for which an

agent acted; in short, bodily movements "are not explained by what the person believed and desired".

(2**)   Intentional explanations of human actions *do* cite reasons for which an agent acted; actions *are* explained by what the person believed and desired.

(3**)   Action (behaviour) *is* a bodily movement.

It should be clear by now that (1**) entails (3*) which in conjunction with (3**) entails (5*) which, in turn, contradicts (2**). Hence, if (1**) and (2**) are true, then a solution to the problem would call to reject (3**). In other words, it would call one to maintain that "it is a mistake to identify action with the external events (e.g., bodily movements) that are the effect of the reasons for which the action is performed. It makes it impossible for reasons to explain the action" (ibid.: 14).

The problem of content epiphenomenalism depends on the view that "an explanation of an action whose explanans is the reason for which the agent acted is a causal explanation" (Alvarez 2007: 105), but the formulation calls for a clarification. For the *explanans* of action is the *reason why* someone acted in the way she did: why, say, John raised his hand. And the reason why is not the *reason for which* or *ground on which* an agent acts even if it is true that the reason why might implicitly or explicitly introduce the latter. So if the fact that John raised his hand is explained by the fact that he wanted to ask a question and knew that raising a hand is what one should do before asking it aloud, then the *explanans* is the fact that he wanted to ask (etc.). But the explanans *includes* reference to the agent, his mental states, as well as to the *contents* of the states. So if content is the *reason for which* an agent behaves in a particular way, then it is wrong to maintain that "explanans is the reason *for* which the agent acted": the explanans includes the reason for which an agent acts, but the explanans itself is the reason *why* – i.e. the fact which explains why – one acts as one does. Furthermore, if the *reason for* is the *content* of a mental state, and the mental states are causes or causal factors of the action, then it is also false

162

that "the reason for which someone performs an action is the cause of the action" (ibid.).

But even if the above considerations are right, they do not show that an explanation of action *is not* a causal explanation. In fact, if what makes an explanation causal is that the explanans – the reason *why* – includes reference to causal factors, then it should be obvious that action explanations *are* causal even if causal factors are not the only factors that would exhaust their explanans.

But according to some philosophers, the idea that *reasons for* are not causes entails that reason explanations or "rationalizations" – i.e. intentional explanations that cite agents' reasons for action – are not a species of causal explanations. In their view, intentional explanations that refer to contents "interpret" actions; to quote Elizabeth Anscombe, they "say something like "See the action in this light"" (Anscombe 1957: 21). On this view, contents which interpret actions are *not* causes. Hence, it is concluded, intentional explanations (or "rationalizations") are *not* a species of causal explanations. But it should be clear by now that the conclusion does not follow.

As Robert Audi notes, if intentional explanations cite reason-states (beliefs, desires, etc.) as well as their contents, and reason-states "play a causal role", then "why not say, then, that while content itself is not a cause, it plays a causal role in explaining action because the reason-states which causally explain action do so in part *by virtue of their content*?" (Audi 1993: 57). In other words, content can have a causal role in explaining action in virtue of the fact that the property of *having content* can be a causally relevant factor for an explanation of action. And if a factor is causally relevant, then an explanation that makes reference to that factor is a causal explanation.

To conclude, the idea that contents (or *reasons for*) are not causes – yet having content *is* a causal factor – is perfectly compatible with the idea that intentional explanations of action are a species of causal explanation. And if explanations of action are causal explanations, then the problem of content epiphenomenalism is still in place.

2. 2. Mental causation and the standard view about action

The idea that mental causation is real is a philosophical way of expressing the idea that the mind can make a causal difference; more specifically, that agents can make causal difference in virtue of possessing mental properties. For example, Pierre Jacob maintains that "there are two commonsense theses about mental causation: a weaker thesis and a stronger thesis".

> According to the weaker thesis, propositional attitudes are causes of intentional action and they are involved in the causal process leading to the formation of new propositional attitudes. (Jacob 1992: 203)

In Jacob's view, the stronger thesis implies that

> tokens of propositional attitudes are causes of intentional action and they are causally involved in the formation of new propositional attitudes in virtue of their contents. On the stronger thesis, not only are mental state tokens causes; in addition, content properties are causally efficacious. (Ibid.)

Two things should be noted at this point. The first concerns terminology. The notion of "causal efficacy" is commonly applied to something that is doing the causal work or brings about change in our universe. But it should be clear that according to the stronger thesis, as well its weaker form, the "causal work" is done not by *content properties*, but by *tokens of propositional attitudes*. Thus, it would be less misleading to describe content properties as being *causally relevant* rather than *causally efficacious*.

Second, it should be indicated that the expression "mental state tokens are causes" is not simply used to stress the fact that it is *tokens* and not *types* that are causes. An equally important implication of this statement is that mental state tokens, whatever they are, fall into the category of causes. However, when scrutinized, the idea that tokens of mental states are causes proves to have important consequences for issues surrounding the phenomenon of mental causation. In order to get a better grasp of the framework that underlies

this seemingly innocent idea one must look at the doctrine of token physicalism.

As Jacob puts it,

> if token physicalism is true […], then the weaker causal thesis is true: if tokens of propositional attitudes are brain state tokens and if brain state tokens can be causes, then so can tokens of propositional attitudes. Token physicalism secures the view (made philosophically respectable by Davidson 1963) that propositional attitudes can be causes or enter causal relations. (Ibid.)

But how exactly does token physicalism secure the view that mental state tokens are causes? More specifically, what must tokens of propositional attitudes *be* for them to fall into the category of *cause*; and what does the category of "cause", when used in the context of token physicalism, delineate?

Eric Marcus notes that Davidson formulated the doctrine of token physicalism to solve a particular problem stemming from a set of seemingly incompatible ideas. Marcus writes as follows:

> Davidson needed a solution to the apparent paradox posed by the conjunction of three ideas: that […] a reduction [of mental types to physical types – MG] is impossible, that the mind is efficacious, and that all causation is physical. The doctrine of token-identity provides such a solution only if the relevant tokens are instances of both mental and physical types – if they have the ontological "thickness" characteristic of particularity. These particulars are causally efficacious (since they are instances of nomological, i.e., physical, kinds), and yet also mental (since they are instances of mental kinds). (Marcus 2009: 216)

Thus, "the question of whether mental states are token-identical to physical states is best understood as a question about the identity of particulars" (ibid.).

According to Marcus' characterization of token physicalism, token states are particulars because only particulars can be "causally efficacious". In other words, the idea of a cause implies the idea of a "particularity"; hence, only

particulars can be causes. Furthermore, token physicalism is based on the "event-view" or the Humean (though not necessarily Hume's) theory of causation, according to which "the only true relata of the causal relation are such things as events and perhaps also states, as itself an event" (Steward 2009: 305). So token states are not *just* particulars; they are *events* which fall into the category of particulars.

But if token states are events that are particulars, then like any particulars they can have two (or more) incompatible and non-equivalent descriptions. Hence, token states *thus understood* can have a mental *as well as* a physical description, and having a physical description ensures that mental states instantiate a natural (or physical) kind. This, in turn, not only satisfies one of the main requirements of ontological naturalism; it also provides an account of how mental causation can play a role in physical causation. As Burge notes,

> the requirement that the physical event-token instantiate a physical kind specifiable in a natural science (physics, chemistry, biology, neurophysiology, and so on) is meant to ensure that the materialist utilize a non-question-begging identification that is not only uncontroversially physical, but plays some role in explanation of physical causation. (Burge 2007: 349-350)

Thus the doctrine of token physicalism not only can explain how the phenomenon of mental causation can be compatible with the idea that all causation is physical; it too can explain how mental states can be causes.

Burge also notes that

> it is common to hold that intentional states and events can be type-identified with functional states and that these states could be realized by a variety of physical states or events. The realizing physical states and events are supposed to be instance- or token-identical with mental states or events. (Ibid.: 354)

But when the doctrine of token physicalism is conjoined with the Humean view that events are the true causes and effects, and is accompanied with the

weak thesis about mental causation, the idea that actions are events seems to follow immediately. The argument is as follows:

(1) Mental states (tokens) are events that cause actions.

(2) Events are the true causes and effects.

(3) Hence, mental state tokens that are events cause events that are actions.

The conclusion of the argument is not equivalent to the idea which gives rises to the problem of content epiphenomenalism – to the idea that actions are bodily movements – but it takes a short step to make it.

> As Jennifer Hornsby notes, the view, according to which mental states are
>
> internal states of people which are the immediate causal ancestors of movements of their bodies […] is certainly held by functionalists. And the functionalists' conception of behaviour may be supposed to recommend itself on the merits of functionalism. (Hornsby 1997: 114)

Furthermore, she indicates that when functionalists speak of behaviour, they generally speak in terms of "bodily movements", "motor responses" or their cognates, and that despite the fact that functionalism is commonly held to inherit the virtues of behaviourism

> the functionalist's notion of behaviour is very much more restrictive than that which some of the behaviourists employed. […] The behaviourist makes allusion to things beyond the agent's body in his specifications of behaviour, but the functionalist does not. And the behaviourist's behavioural items are actions […], whereas the functionalist's behavioural items are apparently not actions, but movements of people's bodies. (Hornsby 1997: 115-116)

One of the main reasons for the functionalist's conception of behaviour is their commitment to giving a reductive analysis of mental concepts. For example, Armstrong reasons as follows:

> We may distinguish between 'physical behaviour', which refers to any merely physical action or passion of the body, and 'behavior proper' which implies relationship to the mind. […] Now, if in our formula

["state of the person apt for bringing about a certain sort of behaviour"] 'behaviour' were to mean 'behaviour proper', then we would be giving an account of mental concepts in terms of a concept that already presupposes mentality, which would be circular. So it is clear that in our formula, 'behaviour' must mean 'physical behaviour'. (Armstrong 1968: 84)

Armstrong's argument is convincing only against the background of *reductive* analysis of mental concepts. But no such assumption is necessary, and conceptual analysis might be illuminating and informative in the relevant ways without taking a reductive form (see Strawson 1992: 17-29). Be that as it may, an important lesson to learn from the above considerations is that there are two ways of speaking about behaviour or bodily movements (or two readings that the notion of "bodily movement" – or its cognates – might take).

Hornsby (1980: Ch. 1) differentiates between two senses of "movement" or "motion" in "bodily movements" or "bodily motion". The distinction corresponds to the transitive and intransitive form of the verb "move". For example, in "A moved B" the verb "move" occurs transitively, whereas in "B moved" it almost always occurs intransitively. Thus, the phrase "a (bodily) movement of B" may either signify an action which consists in A's moving B, in which case it would correspond to the transitive form of the verb; or it may signify B's moving, in which case it would correspond to the intransitive form. Accordingly, it is useful to distinguish between bodily movement$_T$ (where "$_T$" stands for transitive) and bodily movement$_I$ (where "$_I$" stands for intransitive).

If one follows Hornsby and maintains that bodily movements$_T$ are actions, then when functionalists speak of behaviour in terms of bodily movements that are meant to correspond to physical behaviour, the behavior in question is a bodily movement$_I$: in other words, it is not A's moving B, but B's moving; not Mary's moving of her arm (her raising of it), but her arm's moving (its rising).

Now it is not obvious that bodily movements$_T$ are events, and it is even less obvious that they are physical. But if there are physical events at all, then bodily movements$_I$ are among them. Thus, if actions are bodily movements$_I$,

then not only it is true that they are physical events; it is also true that mental state tokens which are events cause events that are bodily movements$_I$. This, in turn, provides an explanation of how mental state tokens that are events can cause *actions*.

The idea that actions are events that are bodily movements$_I$ is commonly known as "the standard story" or "the standard view" about action. The standard view has been primarily inspired by Donald Davidson's seminal article "Actions, Reasons, and Causes" (1963), which was indebted to Curt J. Ducasse's "Explanation, Mechanism, and Teleology" (1925) and Carl G. Hempel's "Rational Action" (1961-62). What is the standard view more specifically?

In Constantine Sandis' view, Ducasse's and Hempel's works

> challenged the 'strong neo-Wittgensteinian current of small red books' and similar-minded works by arguing that (a) actions are events, (b) the reasons that 'rationalize' action are causes of the events in question and (c) the explanation of action makes reference to known, strict, psychophysical laws. (Sandis 2009: 1)

Davidson rejected (c), but his "influential defense of (a) and (b) was so immense that it has come to represent a crucial turning point in philosophical history" (ibid.).

In the present context one can abstract from (b) for two reasons. First, (b) is not essential to the problem of content epiphenomenalism. Second, strictly speaking, (b) is not a claim about action. Thus, if the standard view is to be confined to action, then it can be confined to (a) alone. But (a), however, only tells us that actions are events; it does not tell *what kind* of events are actions or what events *are*.

There is no general consensus on what events are, but the common view is that the events that are actions are "changes in the world". For example, in Berent Enç's book *How We Act: Causes, Reasons, and Intentions* (2003) "actions are defined as changes in the world that are caused by mental events" (Enç 2003: 2) yet the author makes it clear that "this much is not very original.

It follows a tradition that became what used to be called 'the standard view' thanks to a seminal essay by Donald Davidson" (ibid.). For example, G. H. von Wright accepts the above definition of events when arguing *against* the idea that actions are events (von Wright 1963).

So the standard view holds that actions are events which are changes in the world. But what kind of changes in the world *are* actions?

There are several characterizations of the standard story all of which maintain that the kind of events that are actions are bodily movements[12]. For example, Hornsby maintains that "the standard story is sometimes encapsulated in the slogan: 'Beliefs and desires cause actions'" (Hornsby 2004: 2), and that "the standard story of action is "event-based"" (Hornsby 2010: 57). In her view, according to one prominent version of it that is proposed and defended by Michael Smith, "actions are bodily movements that are caused and rationalized by an agent's desire for an end and a belief that moving her body in the relevant way will bring that end about" (Hornsby 2004: 2). Frederick Stoutland proposes a more informative characterization. In his view, the standard story amounts to the thesis that "an action consists of non-intentional bodily movements, the kind neurophysiology describes and explains, which are what a human action is" (Stoutland 2009: 323).

In the philosophical literature, the idea that actions are bodily movements has been formulated in several ways. For example, in Davidson's writings, it amounts to the idea that "if I raise my arm, then my raising my arm and my arm rising are one and the same event" (Davidson 1987b: 103) while Armstrong expresses the same idea by saying that "the rising of my arm is the *raising* of my arm provided that it is part or the whole of a pattern of behaviour that has an objective" (Armstrong 1968: 170).

But the doctrine that actions are bodily movements, however expressed, is only half of the standard story. For even if action *is* a bodily movement, the

---

[12] Henceforth, the notion of "bodily movement" is to be used in the sense of "bodily movement¡" unless indicated otherwise.

latter, according to the standard view, is not *sufficient* for action. As Stoutland notes,

> in order to constitute an action, non-intentional bodily movements must have a special kind of cause – the standard view is that they must be caused (in the right way) by an agent's beliefs and desires – but being so caused does not change their nature as non-intentional bodily movements. (Stoutland 2009: 323)

Hence, a full characterization of the standard story about action amounts to the idea that "actions are events that are identical to movements of the body caused, in a 'non-deviant' way, by a combination of beliefs and so-called 'pro attitudes'" (Sandis 2009: 2).

Now when supplemented by the standard view, the argument presented earlier takes the following form:

(1*)   Mental state tokens cause actions.

(2*)   Mental state tokens are events.

(3*)   Mental state tokens are events that cause actions.

(4*)   Actions are events which are bodily movements.

(5*)   Hence, mental state tokens are events that cause events which are bodily movements.

The standard view is encapsulated in (4*), and the argument shows how, as Dretske notes,

> The Standard View […] fits nicely into a materialistic metaphysics. If beliefs, desires, and fears are physical events of some sort (presumably in the brain of the actor), as they certainly will be on a materialistic view of the world, then there is no problem in understanding how they produce the actions for which they provide (when the reasons are good) a rational justification. They cause them. (Dretske 2009b: 13)

But if the standard view, encapsulated in (4*), is the root of the problem of content epiphenomenalism, what must actions be for contents to be able to explain them?

2. 3. From event- to agent-based account of agency

Dretske defends a view about action which rejects the idea that action is an event that is a bodily movement, and is based on what might be considered to be a *processual account of action*. In his view, action falls under the category of process or occurrence, and not under the category of event as a change in the universe. In that respect, Dretske's view belongs to the minority position – also defended by J. L. Austin, Kenny, von Wright, Alvarez, Hyman and Steward, among a few others – which rejects the doctrine that actions are events. For example, von Wright maintains that "it would not be right, I think, to call acts a kind or species of events. An act *is* not a change in the world. But many acts may quite appropriately be described as the bringing about or effecting (at will) of a change" (von Wright 1963: *f*.35).

According to the processual view, action is not a bodily movement that is caused by a mental state or an agent. Neither it is an event – or a sequence of events – that causes bodily movements. Instead, action is the caus*ing* of bodily movements. When considered carefully, this idea falsifies both (i) the view that action *is* a bodily movement caused by mental states (the standard view; Davidson et al.) or agents (traditional agent causalism; O'Connor 2000; Lowe 2007) and (ii) the view that action is an event which causes bodily movements (Hornsby 1980).

The idea expressed in (i) is mistaken because if action is a causing of bodily movement, and causing of bodily movement cannot be identical to the movement itself, then action cannot be a bodily movement. The idea (ii) is false because if action is a causing, then it is not a cause, and if it is not a cause, then it is not an event that *is* a cause of bodily movement. Hyman explains why a caus*ing* cannot be a cause in the following way:

> Where '*a*' and '*b*' are names of particulars (of whatever categories or kinds) and '*ϕ*' is a causative verb or verb phrase, 'a's *ϕ*ing of *b*' does not refer (except by accident) to an object or event that *ϕ*s *b*. In every case,

it refers to the ϕing not the ϕer: the killing not the one that kills, the raising not the one that raises, and in general the causing and not the one that causes. […] In short, a caus*ing* is not a cause. (Hyman 2015: 57)

Now if the idea that action is a caus*ing* of a bodily movement can be defended, then it would falsify (4*), thus eliminating the root of the problem of content epiphenomenalism. But if action is a causing of a movement rather than a bodily movement that is being caused, what kind of causing it is?

There are two opposing views on this question, which are based on different accounts of agency: the event-based account of the type Dretske defends, and the agent-based account of the type defended by Alvarez, Hyman and Steward.

Dretske's view is that "behavior is not the bodily movements that internal events cause, but the causing of these movements by internal events" (Dretske 2009b: 21). Thus, "moving your arms and legs (behavior) is not the same as the movements of the arms and legs, but rather an internal event causing the arms and legs to move" (ibid.).

According to Dretske, behavior that is an action is a process that includes, as a component, a mental state (a "triggering cause") which by virtue of its content properties (a "structuring cause") causes a bodily movement. More formally, where "a" is a triggering cause having a structuring cause, and "b" is an event that is bodily movement, Dretske's view amounts to the idea that action is a process of *a causing b*. Hence, even if neurophysiological properties are causally relevant and sufficient for an explanation of why *b* has occurred, this does not show that content properties of *a* cannot be causally relevant for an explanation of why an *action* occurred. For action, in Dretske's view, is not *b*, but *a's causing b*.

But Dretske's theory of action fails at least for two reasons.

The first concerns the theory's implications for perception. For if action is an internal event of causing bodily movement, then actions cannot be perceived: internal events that cause bodily movements occur *inside* the agent

and what occurs inside cannot be perceived[13]. What *can* be perceived, if the account is right, are only *parts* of actions: more specifically, a part which is caused by an internal event – a bodily movement. But if that is right, then it is false that we see agents acting. What is true, instead, is that we see only parts of their actions, or perhaps their consequences.

The second objection to Dretske's theory takes a more complex form. Dretske defines behaviour as an internal event causing bodily movement, where an internal event (a "triggering cause") is mental state token. If that is right, then Dretske's picture of action is based on the doctrine of token physicalism. But Dretske not only assumes that mental state tokens are internal events; he also holds that mental state tokens cause *bodily movements*. In other words, Dretske assumes that (5*) is true. But what are his reasons for holding the idea expressed in (5*)?

It does not seem that the doctrine of token physicalism is capable of supporting (5*). For neither the idea that mental states are token identical to physical states nor the idea that mental state tokens are internal events entail that these tokens cause bodily movements, even if it is true that they *might* or *do* cause *other events*. Nothing in the doctrine of token physicalism seems to rule out the idea that the token states (*qua* events) cause events which are *not* bodily movements. Furthermore, the view that bodily movements of the relevant sort have neural causes – are caused by neural events – does *not* give support to the idea that the neural events are the relevant token states.

As Jacob maintains, token physicalism "secures the view […] that propositional attitudes can be causes or enter causal relations" (Jacob 1992: 203). But the latter view is not equivalent to saying that it gives support to the view that mental states are causes of actions: the weaker thesis of mental causation is part of commonsense, not part of the theory of token physicalism. And the weaker thesis does not show that mental states are causes of bodily movements.

---

[13] Conditions when internal events can be perceived – say, in a laboratory – are irrelevant for the argument that is restricted to perception in non-artificial settings.

If the above considerations are right, then token physicalism does not secure the view that mental states (their tokens) cause bodily movements. Instead, the idea that mental states cause bodily movements *presupposes* the weaker thesis about mental causation which describes mental states *as causes of action*. Moreover, as the presented arguments show, the view that mental states cause bodily movements *rests* on the problematic assumption that actions *are* bodily movements. But if that is the case, then Dretske's theory which defines action as causing of internal events that cause bodily movements seems to presuppose the incompatible idea that actions are bodily movements.

Dretske's processual theory of action aims to "reduce agency to the case of an event being a cause" (Davidson 1980: 128), which shows that Dretske's theory, as like event-based account of agency, is linked to the Humean view of causation. The general idea is to characterize agency in terms of concatenation of events, thus reducing agency to a sequence of events that cause each other. This, in turn, rests on the premise that causation by agents can (and should) be reduced to causation by events.

But not all philosophical theories of agency are based on the idea that all causation is event-causation. As Hornsby notes,

> in recent years, there has been a rise in a neo-Aristotelian, anti-Humean approaches to causation. The initial anti-Humean thought may be that the very various phenomena which attract the label "causation" cannot all be brought under the head of "cause" if this is understood as a relation between events. More specifically, the suggestion may be that a different understanding is required in order that *agency* – a phenomenon in which something *acts* – should be recognized as the causal phenomenon it is. (Hornsby 2015: 130)

The kind of theories that can be described as following the anti-Humean thought in the above sense have traditionally been subsumed under the rubric of "agent causalism". Proponents of the theories of this kind generally disagree with the Davidsonian claim that agency can be reduced to "an event being the

cause"; in other words, they reject the core element of the standard story of action. As Hyman notes,

> partisans of agent-causation argue that event-causation alone cannot amount to any more than regularity or counterfactual dependence (isn't this the main lesson of the analysis of causation in Hume's *Treatise*?), and that the idea of an origin or source, which is essential to the idea of causation, cannot be captured in this way (isn't *this* the main lesson of the long history of inadequate Humean theories of causation?). (Hyman 2015: 41)

Now surely the partisans of agent-causation do not deny the existence of event-causation; what they maintain is that the idea of agent causation is necessary in order to capture "the idea of an origin or source", which they take to be necessary to the idea of causation itself. As Hacker puts it,

> in reaction to such Humean and neo-Humean analyses of causation there have been various, very different attempts, by Collingwood, Gasking, Hart and Honoré, Anscombe, von Wright and others, to link the concept of causation neither to regular sequences of events nor to nomologically connected events simpliciter, but rather, directly or indirectly, to human agency or to human action and to the possibility of manipulation – of bringing about events through human intervention in the normal course of events. (Hacker 2007: 60)

It would be a mistake to conclude that the above idea is incompatible with the phenomenon of event-causation. Quite to the contrary: partisans of agent-causation hold that event-causation co-exists with agent- or substance-causation, yet it is the latter that captures the idea of an *origin* of causal intervention.

But there is more than one way that an agent-based account of agency can be developed. For example, according to the so-called traditional agent causalists, agency is a case of agents causing events that are their actions: to act, according to this view, is for agents to cause their actions.

There are several objections to the traditional agent causalism, but for present purposes it should suffice to note it rests on the standard view about action – that is, it assumes the doctrine that actions are bodily movements. Consequently, the traditional agent causalism is committed to the problem of content epiphenomenalism. Thus, if a theory of action is to be based on the idea of agency that is agent- as opposed to event-based, it must be a kind of theory that would *not* assume (or be committed to) the problematic doctrine that action is a bodily movement.

A natural alternative would be to retain the idea that agency is to be characterized by reference to the idea of agent causation while denying that what an agent causes is a piece of action. So one might agree with Dretske that actions are caus*ings* of events that are bodily movements yet reject the idea that agency can be reduced to event-causation. One can follow Alvarez, Hyman, and Steward, and define action as, roughly, an exercise of a two-way causal power; more specifically, one might hold that agency "is an exercise of bodily control on the part of its agent – the bringing about of a bodily movement or change by that agent by means of the exercise of its two-way power to do so" (Steward 2009: 308). According to this view, then, "to act is to exercise a causal power – to cause, bring about or effect an event" (Alvarez & Hyman 1998: 233) where "the exercise of a causal power is neither an event, nor the relation between agent and event that it entails" (ibid.).

Now if one thinks of agency and action as a causing of an event that involves an exercise of a two-way causal power done by agents, then one can block the idea that content properties are irrelevant to the explanation of action even if neurophysiological properties of internal events are sufficient for causal explanations of bodily movements. For even if it is true that content properties are irrelevant to explanations of why a bodily movement occurred (and they *are* irrelevant if "the neurophysiological story" is not deficient or incomplete), they might still be relevant (and they *are* relevant if actions are to be explained by reference to them) to an explanation of why an agent *caused* such a movement to occur. In short, content properties are irrelevant to the explanans

of why John's *arm raised*, but are relevant to the explanans of why *John has raised his arm*, and the two are not the same, if the above view of action is right.

Contrary to the standard view of action, the agent-based account of agency is not only based on the idea of substance-causation; it too rests on the ontology of powers. For according to this view, to act is not for an internal event to cause another event in virtue of its properties; rather, to act is for a substance (and agent) to cause an event by exercising a causal power. Furthermore, on this view, instantiations of content properties (or mental properties more generally) do not *cause* actions or bodily movements. So they are not, as Hutto notes, "causally efficacious mental types – properties that when appropriately instantiated productively and mechanically bring about actions in the way supposed by traditional analytic functionalists" (Hutto 2013: 67). On the agent-based account of causation, mental properties are causal factors which are causally relevant for – though do not "mechanically bring about" – the occurrence (and explanation) of action.

Some philosophers (for example, Kim 2000: 75; 2005: 22; 2007: 236) think that if mental properties are only causally relevant for explanations of action, then reason explanations are not genuine causal explanations. But the plausibility of this objection depends on the premise that mental properties *must* be of a productive kind (in the sense explained), which, in turn, rests on the functionalist's assumption "that a certain narrowly based, input/output version of common sense functionalism that defines mental states as internal causes is a bona fide characterization of our folk commitments" (Hutto 2013: 67). In other words, only if mental states are construed as internal causes of actions can mental properties be considered to be of a productive sort. But the idea that mental states must be internal causes of actions – the token-identity theory of mental states – goes together with the doctrines that actions are bodily movements as well as with the event-based account of agency. Hence, if the latter philosophical views are rejected, then the idea that mental states cause actions does not, by itself, support the idea that mental states are (token-

identical to) internal causes of bodily movements, and thus does not imply that mental properties (or their instantiations) must be causes in a productive or mechanistic way. In fact, if the Aristotelian account of causation is adopted, then properties (or their instantiations) are not the kind of objects that fall into the category of causes (or objects that *cause*). Rather, according to this view, properties – both physical and mental – are causally relevant factors which play an ineliminable *causal* role in the causal processes of real producers of effects – that is, of substances and of events.

Naturally, the processual and agent-based conception of action can hardly be described in terms of categories that are implicit in the event-based Humean view of causation. As Steward writes,

> an exercise of power has to be conceptualized, on the dominant Humean view of causation [...] as itself an event, and therefore as something which must itself be caused by prior events and states. The exercise cannot just be thought of, unreduced, as just *that*, an exercise of a power by a substance, for the notion of an exercise of power by a substance is deemed unintelligible, unless is can be shown to be equivalent to the notion of an event occurring *within* that substance producing an effect. (Steward 2009: 205)

Consequently, if the Humean view about causation is what guides one's theory of agency and action, it is tempting to "reduce the concept of an exercise of power by S to the concept of a causal relation between causes which are thought of as events and states – such as beliefs, desires, intentions, choices and so on and an event-like effect, a movement of some sort" (Steward 2009: 306).

But even if it is true that the concept of an exercise of a causal power cannot be reduced to the concept of causal relations holding between events, it might still be compatible with the idea that "agents can exercise causal powers by dint of events" (Hyman 2015: 41). In fact, if Hyman is right, then it seems that both agent- and event-causation are interdependent. As he notes,

179

we can explain how a man moves his hand by describing the events inside his body that cause the motion of his hand: the release of neurotrasmitters, the contraction of muscles, and so on. But these events are caused by nerve cells and muscle fibres, which are agents themselves, with their own causal powers. These powers in turn are activated by events caused by more minute agents. And so on, until, below the nanoscale, substance, power, action all dissolve, and causation is 'swallowed up in mathematics', in other words pure regularities, or patterns of functional dependence between events. (Hyman 2015: 41-42)

But the agent-based account of agency as well as the idea of agent causation more generally is not free of criticism. Two of them shall be considered below.

It is sometimes argued that the agent-based conception of agency is misguided because it is incompatible with the idea that only events can cause events. Now the idea that *only* events can cause events must be distinguished from the weaker idea that every event that has a cause is caused by another event, which might be true. But the latter proposition does not entail that agents *cannot* cause the same events that are caused by other events. And neither does the weaker claim give support the view that *only* events can cause events. Surely, the thought that only events can cause stems from the Humean modes of thought, and it is true if the Humean event-based account of causation is true. But if the idea of agent causation is made intelligible and is shown to be compatible with the proposition that event-causation is necessary for the exercise of causal powers of agents, then the weaker idea might be all that is needed for an adequate account of how causation works.

The second objection consists in the idea that agents cannot cause events for if events are caused by other events, then on the pain of over-determination those same events cannot be *also* caused by agents. But the objection is not decisive if, again, it is not assumed that causation should be characterized only by reference to causation by events. One should start by distinguishing between two different ways causation can manifest itself: one as it is

manifested in event-causation, another when it is manifested in substance- or agent-causation. In the former case, the category of "cause" applies to a two-place extensional relation that holds between two items of the same categorical kind. But if agent causation is based on the idea of an exercise of causal power, and "the exercise of a causal power is neither an event, *nor the relation between agent and event that it entails*" (italics – M.G.) (Alvarez & Hyman 1998: 233), then caus*ing* of an event by an agent (and so the notion of "cause") does *not* apply to a relation. As Hornsby notes,

> Neo-Aristotelians do not treat *cause* as everywhere a relation – neither as a relation between two events, nor between two objects, nor between an object and an event. Neo-Aristotelians find fault with the empiricists' treatment of dispositional properties as analyzable away in favour of counterfactual conditionals which introduce relations between events. […] Thus they defend a metaphysics in which a substance ontology belongs, and to which such notions as powers, capacities, liabilities are central. (Hornsby 2015: 131-132)

So according to the anti-Humean neo-Aristotelian theory of causation, the notion of "cause" when used in "*S* causes *b*" (as in "John caused the crash") does *not* indicate the same thing as when it is used in "*a* causes *b*" (as in "Neural events in John's brain caused the crash"). So it might be true that if *b* is caused by *a*, then *b* cannot be *also* caused by *S*. But if it *is* true, then the notion of "cause" in both the antecedent and consequent of the conditional must be used to indicate the same kind of thing: more specifically, the same kind of causation in process. However, if the notion of "cause" does *not* carve out the same kind of thing in both cases – that is, if agent- or substance-causation is of a different kind than event-causation – then the conclusion does not follow, and the objection fails.

If the considerations in this chapter are right, then the problem of content epiphenomenalism stems from the problematic doctrine that defines actions as bodily movements, thus creating a common explanandum for intentional and neurophysiological explanations. The roots of the doctrine lie in the event-

based account of agency that is implicit in the orthodox forms of functionalism as well as in the Humean modes of thought about causation more generally. An alternative view of agency that rests on the idea of agent causation and defines action in terms of agents' causing of events by means of an exercise of causal powers can provide a plausible position that could account for the intuition that contents can be relevant for explanations of actions given that neurophysiological explanations of bodily movements are not deficient or incomplete.

Conclusions

1. The results of the analysis show that the general problem of intentionality, which aims at the questions of *what it is* or *what must be the case* for something (a mental state, a linguistic expression, a gesture, etc.) to have a particular content, calls for a constitutive account of intentionality that aims to specify its essence or nature. Furthermore, taking into account the grammatical structure of intentional attitude reports, the problem of intentionality should be interpreted as encompassing three distinct yet closely interconnected groups of questions that are distinguished by reference to their subject-matter and characterized by a set of questions concerning the indicated object of inquiry. The standard form of semantic naturalism is additionally committed to the idea that the relevant kind of theory of intentionality ought to be reductive and construed in terms of some natural science.

2. According to the prevalent conception, a constitutive account of some phenomenon aiming to specify its nature ought to be understood as calling for a specification of metaphysically necessary and sufficient conditions that ought to be substantiated by analyses. Taking into the account the critique of the idea of concept-independent essence and irreducible, explanatorily basic idea of metaphysical necessity, it is being suggested that the character of constitutive accounts, as well as the character of the modality and substantiation they involve, should be interpreted on the conceptualist lines: essence is conceptual (what is essential for X follows from the concept of X) as is the kind of modality implicit in constitutive accounts (the necessity of the constitutive claim follows from the idea that it is a conceptual claim, explicating or defining the concept of X), and substantiation by analysis should be taken to call for substantiation by conceptual analysis, which might take a variety of forms (i.e. not necessarily the specification of individually necessary and jointly sufficient conditions) and be intertwined with empirical inquiry. Finally, in as much as the essence of the phenomenon should be distinguished

from its causal-enabling preconditions, or sufficient conditions more generally, constitutive accounts should be distinguished from causal-enabling explanations, as well as explanations that are construed on the basis of categories of *supervenience* or *grounding*.

3. The analysis of the general framework of semantic naturalism shows that the predominant versions of naturalistic theories of semantics belong to the type of causal theories of mental content (and more generally, to causal theories of mind), presuppose the standard version of semantic externalism, and are based on the functionalist and representational theories of mind. The fact that they belong to the group of causal theories of mental content is in part determined by their externalist character, which, in turn, is intertwined with their commitment to the categorical structure of functionalism and the representational theory of mind and concepts. Furthermore, the analysis shows that the commitment to functionalism and representationalism, which is accompanied by the standard version of semantic externalism, amounts to a bifurcationist conception of intentional attitudes.

4. A theory of intentionality that is based on the categories of *ability* and *power* is a preferable alternative to the predominant naturalistic accounts of intentionality, which prove to be essentially Cartesian and bifurcationist, and are based on the categories of *internal state*, *mental representation* and *causal relations* and/or *biological functions* determining the semantic properties of representations. The alternative, ability-based approach possesses the following theoretical virtues:

    i. It can explain the problems of semantic normativity, fine-grainedness of content, self-knowledge and concept shareability that the standard naturalistic theories face and have difficulty accounting for;

    ii. It is compatible with a version of semantic externalism that (a) does not imply the view that external factors unknown to the subject directly determine the intentional content, (b) represents the external environmental conditions as a relevant contextual factor for determining content, and (c) therefore does not have the epistemically problematic

implication that is generally taken to follow from the standard version of externalism, namely – a denial of first-person authority;

iii. It is compatible with, and so can be construed on the basis of Fregean *Sinn* yet this need not imply that a theory is committed to semantic internalism;

iv. Despite the fact that it does not meet the requirements of scientific naturalism – it is neither reductive, nor construed in terms of the vocabulary or ontological commitments of some natural science (such as physics, chemistry or biology) – it is naturalistic enough and thus is a preferable naturalistic account of intentionality.

5. The analysis of the problem of content epiphenomenalism, which is a daunting problem for naturalistic theories of mind and action, indicates that the problem is tied to the standard view of action, which is guided by the ideal to propose a reductive account of agency in terms of event-causation (to define action in terms of concatenation of events), is based on a particularist ontology of mind implied by token physicalism and rests on the Humean theory of causation. After a critical analysis of several alternative accounts of action, it is suggested that the problem could be accounted for if an alternative theory of action, based on the category of agent causation, Aristotelian theory of action and power-based ontology of mind is admitted. An alternative to the standard view of action that is being proposed defines action not in terms of an event that is caused by token states, but in terms of a process during which an agent causes an event. Hence, by rejecting the premise following from the standard view, which implies that action is an event that is a bodily movement (in the intransitive – "A moved" as opposed to "A moved B" – sense), one rejects the main root of the problem of content epiphenomenalism: namely, the idea that action explanation, in which reference to intentional content plays an essential role, is an explanation of events that are bodily movements (in the intransitive sense).

Bibliography

Adams, F. 2003. Thoughts and Their Contents: Naturalized Semantics. In: S. Stich and T. Warfield, eds. *The Blackwell Guide to Philosophy of Mind*, Oxford: Basic Blackwell, p. 143-171.

Alvarez, M. 2007. The Causalist/Anti-Causalist Debate in the Theory of Action: What It Is and Why It Matters. In: A. Leist, ed. *Action in Context*, Berlin: Walter de Gruyter, p. 103-124.

Alvarez, M., Hyman, J. 1998. Agents and Their Actions. *Philosophy* 73(2): 219-245.

Ammereller, E. 2001. Wittgenstein on Intentionality. In: H. J. Glock, ed. *Wittgenstein: A Critical Reader*. Oxford: Wiley-Blackwell, p. 59-93.

Anscombe, E. 1957. *Intention*. Cambridge, MA: Harvard University Press.

Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.

Armstrong, D. M. 1981. *The Nature of Mind and Other Essays*. Cornell: Cornell University Press.

Armstrong, D. M. 1997. *A World of States of Affairs*. Cambridge, MA: Cambridge University Press.

Audi, R. 1993. *Action, Intention, and Reason*. Ithaca, NY: Cornell University Press.

Baker, L. R. 1995. *Explaining Attitudes: A Practical Approach to the Mind*. Cambridge, MA: Cambridge University Press.

Baker, L. R. 1998. Cognitive Suicide. In: R. H. Grimm and D. D. Merill, eds. *Contents of Thought-Arizona Colloquium in Cognition*. Tuscon: The University of Arizona Press, p. 1-30.

Bell, D. 1979. *Frege's Theory of Judgement*. Oxford: Clarendon Press.

Ben-Yami, H. 2001. The Semantics of Kind Terms. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 102(2): 155-184.

Bilgrami, A. 1992. *Belief and Meaning: The Unity and Locality of Mental Content.* Oxford: Wiley-Blackwell.

Block, N. 1978. Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science* 9: 261-325.

Block, N. 1981. Psychologism and Behaviorism. *Philosophical Review* 90(1): 5-43.

Boghossian, P. 1989. The Rule-Following Considerations. *Mind*, New Series 98(392): 507-549.

Brentano, F. 1995. *Psychology from an Empirical Standpoint*. New York: Routledge.

Bridges, J. 2006. Does Informational Semantics Commit Euthyphro's Fallacy? *Noûs* 40(3): 522-547.

Burge, T. 1977. Belief de re. *The Journal of Philosophy* 74(6): 338-362.

Burge, T. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* 4: 73-121.

Burge, T. 1986. Individualism and Psychology. *The Philosophical Review* 95(1): 3-45.

Burge, T. 1988. Individualism and Self-Knowledge. *The Journal of Philosophy* 85: 649-663.

Burge, T. 2003. Phenomenality and Reference: Reply to Loar. In: M. Hahn and B. Ramberg, eds. *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, MA: MIT Press, p. 435-450.

Burge, T. 2007. *Foundations of Mind*. New York: Oxford University Press.

Burge, T. 2010. *Origins of Objectivity*. New York: Oxford University Press.

Cain, M. J. 2002. *Fodor: Language, Mind, and Philosophy*. Cambridge: Polity Press.

Campbell, J. 2011. Intelligible Causation. In: J. Bridges, N. Kolodny, W. Wong, eds. *The Possibility of Philosophical Understanding: Reflections on the Thought of Barry Stroud*. New York: Oxford University Press, p. 232-249.

Caston, V. 2007. Intentionality in Ancient Philosophy. *Stanford Encyclopedia of Philosophy*, [online] available at: http://plato.stanford.edu/entries/intentionality-ancient/ [accessed 2015 03 26].

Child, W. 1994. *Causality, Interpretation, and the Mind*. Oxford: Oxford University Press.

Chomsky, N. 1959. A Review of B. F. Skinner's Verbal Behavior. *Language* (35)1: 26-58.

Craig, E. 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Oxford University Press.

Crane, T. 2013. *The Objects of Thought*. Oxford: Oxford University Press.

Dagys, J. 2006a. Dekartiškas dualizmo įrodymas šiuolaikinėje sąmonės filosofijoje. *Problemos* 69: 95-103.

Dagys, J. 2006b. Funkcionalizmas sąmonės filosofijoje: metodologija ar ontologija. *Problemos* 70: 113-125.

Dagys, J. 2006c. *Reduktyvistinės sąmonės teorijos šiuolaikinėse filosofijos kontroversijose*, daktaro disertacija: humanitariniai mokslai, filosofija (01 H). Vilniaus universitetas.

Dagys, J. 2007. Pareitis, fizikalistinė redukcija ir perkeistos kokybės. *Problemos* 71: 108-115.

Dagys, J. 2008. The Mind-Body Problem and Libertarian Freedom of the Will. *Problemos. Supplement*: 55-62.

Dagys, J. 2012. Šiuolaikinė psichofizinio reduktyvizmo ir antireduktyvizmo polemika ir jos metodologinės išdavos. *Problemos. Supplement*: 22-36.

Dagys, J., Dranseika, V., Grenda V., Vasilionytė, I. 2014. Reductionism in Contemporary Analytical Philosophy. *Problemos* 85: 141-152.

Dainton, B., Robinson, H. 2014. Analytic Philosophy of Mind. In: B. Dainton and H. Robinson, eds. *The Bloomsbury Companion to Analytic Philosophy*. London: Bloomsbury Publishing, p. 128-137.

Daly, C. 2012. Skepticism about Grounding. In: F. Correia and B. Schnieder,

eds. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press, p. 81-100.

Davidson, D. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy* 60: 685-700.

Davidson, D. 1980. *Essays on Actions and Events*. New York: Oxford University Press.

Davidson, D. 1987a. Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, 61: 441-458.

Davidson, D. 1987b. Problems in the Explanation of Action. In: P. Petit, R. Silvan, and J. Norman, eds. *Metaphysics and Morality: Essays in Honour of J. J. C. Smart*. Oxford: Basil Blackwell, p. 35-49.

Davidson, D. 1991. Epistemology Externalized. *Dialectica* 45: 191-202.

Davidson, D. 1997. Seeing Through Language. In: J. Preston, ed. *Thought and Language*. Cambridge: Cambridge University Press, p. 268-286.

Davidson, D. 2001. *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press

Davidson, D. 2004. *Problems of Rationality*. Oxford: Oxford University Press.

Davidson, D. 2005. *Truth, Language, and History*. Oxford: Oxford University press.

Davies, M. 2005. Cognitive Science. In: F. Jackson and M. Smith, eds. *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press, p. 358-396.

Davis, S. 1983. *Causal Theories of Mind: Action, Knowledge, Memory, Perception and Reference*. Berlin: Walter de Gruyter.

De Caro, M., Macarthur, D. eds. 2004. Introduction. In: *Naturalism in Question*. Cambridge, MA: Harvard University Press, p. 1-20.

DeGrazia, D. 1996. *Taking Animals Seriously*. Cambridge: Cambridge University Press.

Degutis, A. 1984. *Kalba, mąstymas ir tikrovė*. Vilnius: Mintis.

Degutis, A. 2007. *Kalba, pažinimas ir veiksmas (rinktiniai straipsniai*

analitinės filosofijos temomis). Vilnius: Versus Aureus.

Dennett, D. 1969. *Content and Consciousness*. New York: Routledge.

deRosset, L. 2013. Grounding Explanations. *Philosophers' Imprint* 13(7): 1-26.

Donnellan, K. 1970. Proper Names and Identifying Descriptions. *Synthese* 21(3/4): 335-358.

Dranseika, V. 2010. Veiksmų atpažinimas kaip *virtus dormitiva*. *Problemos* 77: 122-128.

Dranseika, V. 2012. Reductionism and Action Predicates. *Problemos. Supplement*: 51-62.

Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F. 1996. Phenomenal Externalism or If Meanings Ain't in the Head, Where are Qualia? *Philosophical Issues* 7: 143-158.

Dretske, F. 2000. *Perception, Knowledge, and Belief: Selected Essays.* Cambridge: Cambridge University Press.

Dretske, F. 2009a. Information-theoretic semantics. In: B. McLaughlin, A. Beckermann, S. Walter, eds. *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press, p. 381-394.

Dretske, F. 2009b. What Must Actions Be for Reasons to Explain Them? In: C. Sandis, ed. *New Essays on the Explanation of Action*. London: Palgrave Macmillan, p. 13-21.

Ducasse, C. J. 1925. Explanation, Mechanism, and Teleology. *The Journal of Philosophy* 22(6): 150-155.

Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.

Dummett, M. 1993. *The Seas of Language*. Oxford: Clarendon Press.

Dupré, J. 1996. The Mental Lives of Nonhuman Animals. In: M. Bekoff and D. Jamieson, eds. *Readings in Animal Psychology*. Cambridge, MA: Press,

p. 323-336.

Enç, B. 2003. *How We Act: Causes, Reasons, and Intentions*. Oxford: Clarendon Press.

Epstein, B. 2015. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Clarendon Press.

Fine, K. 1994. Essence and Modality. *Philosophical Perspectives* 8: 1-16.

Fine, K. 2012. Guide to Ground. In: F. Correia and B. Schnieder, eds. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge, MA: Cambridge University Press, p. 37-80.

Fodor, J. A. 1975. *The Language of Thought*. MA: Harvard University Press.

Fodor, J. A. 1984. Semantics, Wisconsin Style. *Synthese* 59(3): 231-250.

Fodor, J. A. 1985. Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum. *Mind*, New Series 94(373): 76-100.

Fodor, J. A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. 1990. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fodor, J. A. 1994. *The Elm and the Expert*. Cambridge, MA: The MIT Press.

Fodor, J. A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Fodor, J. A. 2001. Language, Thought and Compositionality. *Mind and Language* 16(1): 1-15.

Fodor, J. A. 2003. *Hume Variations*. Oxford: Clarendon Press.

Frege, G. 1993. Sense and Reference. In: A. W. Moore, ed. *Meaning and Reference*. Oxford: Oxford University Press, p. 23-42.

Geach, P. 1957. *Mental Acts: Their Content and their Objects*. London: Routledge & Kegan Paul.

Gettier, E. 1963. Is Justified True Belief Knowledge? *Analysis* 23: 121-123.

Glock, H. J. 2001. Intentionality and Language. *Language and Communication*

21(2): 105-118.

Glock, H. J. 2003. *Quine and Davidson on Language, Thought, and Reality*. Cambridge: Cambridge University Press.

Glock, H. J. 2008. *What is Analytic Philosophy?* Cambridge: Cambridge University Press.

Glock, H. J. 2010. Concepts: Between the Subjective and the Objective. In: J. Cottingham and P. Hacker, eds. *Mind, Method, and Morality: Essays in Honour of Anthony Kenny*. Oxford: Oxford University Press, p. 306-329.

Glock, H. J. 2012. What is a Theory of Meaning? Just when you thought conceptual analysis was dead… . *Cahiers Ferdinand de Saussure* 65: 51-79.

Glock, H. J. *Forthcoming*. *Animal Minds: Conceptual Problems*. Book Manuscript.

Glock, H. J., Preston, J. 1995. Externalism and First-Person Authority. *The Monist* 78(4): 515-33.

Glüer, K. 2014. Theories of Meaning and Truth Conditions. In: M. Garcia-Carpintero and M. Kölbel, eds. *The Bloomsbury Companion to the Philosophy of Language*. New York: Bloomsbury Academic, p. 84-105.

Goldberg, S. 2007. *Internalism and Externalism in Semantics and Epistemology*. Oxford: Oxford University Press.

Greenberg, M. 2005. A New Map of Theories of Mental Content: Constitutive Accounts and Normative Theories. UCLA Public Law & Legal Theory Series, [online] available at: http://escholarship.org/uc/item/2k58b229 [accessed 2015 11 04].

Grice, H. P. 1957. Meaning. *The Philosophical Review* 66: 377-88.

Grice, H. P. 1961. The Causal Theory of Perception. *Proceedings of the Aristotelian Society*. *Supplement* 35: 121-168.

Grice, H. P. 1969. Utterer's Meaning and Intentions. *The Philosophical Review* 68: 147-77.

Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard

University Press.

Hacker, P. M. S. 2007. *Human Nature: The Categorial Framework*. Oxford: Blackwell Publishing.

Hacker, P. M. S. 2011. A Plague on both your –isms. *American Philosophical Quarterly* 48(2): 97-111.

Hacker, P. M. S. 2013. *The Intellectual Powers: A Study of Human Nature*. Oxford: Wiley-Blackwell.

Hanfling, O. 2000. *Philosophy and Ordinary Language.* London: Routledge.

Hare, R. M. 1952. *The Language of Morals*. Oxford: Oxford University Press.

Hempel, C. G. 1961-1962. Rational Action. *Proceedings and Addresses of the American Philosophical Association* 35: 5-23.

Hofweber, T. 2009. Ambitious, Yet Modest, Metaphysics. In: D. Chalmers, D. Manley, and R. Wasserman, eds. *Metametaphysics*, Oxford: Oxford University Press, p. 260-289.

Holton, R. 2000. Review of John McDowell's *Mind, Value, and Reality*, [online] available at: http://web.mit.edu/holton/www/pubs/McDowell.pdf [accessed 2016 11 24].

Horgan, T., Timmons, M. 1993. Metaphysical Naturalism, Semantic Normativity, and Meta-Semantic Irrealism. *Philosophical Issues (Naturalism and Normativity)* 4: 180-204.

Hornsby, J. 1980. *Actions*. London: Routledge & Kegan Paul.

Hornsby, J. 1997. *Simple Mindedness: In Defense of Naive Naturalism in the Philosophy of Mind*. Cambridge: Harvard University Press.

Hornsby, J. 2000. Personal and Sub-Personal: a Defence of Dennett's Early Distinction. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 3(1): 6-24.

Hornsby, J. 2004. Agency and Actions. In: J. Hyman and H. Steward, eds. *Agency and Action*. Cambridge: Cambridge University Press, p. 1-24.

Hornsby, J. 2010. The Standard Story of Action: An Exchange (2). In: J. H. Aguilar and A. A. Buckareff, eds. *Causing Human Actions: New*

Perspectives on the Causal Theory of Action. Cambridge, MA: MIT Press, p. 57-69.

Hornsby, J. 2015. Causality and "the mental". *Humana Mente* 29: 125-140.

Houdé, O. ed. 2004. *Dictionary of Cognitive Science: Neuroscience, Psychology, Artificial Intelligence, Linguistics, and Philosophy*. Hove: Psychology Press.

Hutchinson, P. 2008. *Shame and Philosophy: An Investigation in the Philosophy of Emotions and Ethics*. London: London: Palgrave Macmillan.

Hutto, D. 2008. *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.

Hutto, D. 2013. Still a Cause for Concern: Reasons, Causes and Explanations. In: A. Laitinen, C. Sandis, and G. D'Oro, eds. *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*. London: Palgrave Macmillan, p. 49-74.

Hyman, J. 1992. The Causal Theory of Perception. *The Philosophical Quarterly* 42(168): 277-296.

Hyman, J. 1993. Vision, Causation and Occlusion. *The Philosophical Quarterly* 43(171): 210-214.

Hyman, J. 2006a. Knowledge and Evidence. *Mind* 115(460): 891-916.

Hyman, J. 2006b. *The Objective Eye: Color, Form, and Reality in the Theory of Art*. Chicago: University of Chicago Press.

Hyman, J. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

Hyman, J. 2017. Review of *Strange Tools: Art and Human Nature*, by Alva Nöe. *Mind* 126(501): 304-309.

Jackson, F. 1998. Identity Theory of Mind. In: E. Craig, ed. *Routledge Encyclopedia of Philosophy*. New York: Routledge, p. 395-399.

Jacob, P. 1991. Review Work: *Mental Content* by Colin McGinn. *The Journal of Philosophy* 88(12): 723-728.

Jacob, P. 1992. Externalism and Mental Causation. *Proceedings of the Aristotelian Society*, New Series 92: 203-219.

Japertas, M. 1998. *Reikšmės ir tiesos santykio problema šiuolaikinėse reikšmės teorijose*, daktaro disertacija: humanitariniai mokslai, filosofija (01 H). Vilniaus universitetas.

Japertas, M. 2001. Antirealistiniai argumentai filosofijoje. *Problemos* 59: 61-76

Japertas, M. 2005. Frege's įnašas į reikšmės sampratą. *Problemos* 68: 82-91.

Japertas, M. 2007. Empirizmo dogmų kritika: užmojai ir ribos. *Problemos* 72: 104-114.

Japertas, M. 2012. Supratimo epistemologijos keblumai: McDowellas prieš Dummettą. *Problemos. Supplement:* 89-98.

Kaplan, D. 1989. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In: J. Almong, J. Perry, and H. Wettstein, eds. *Themes from Kaplan*. New York: Oxford University Press, p. 481-564.

Katz, J. 1990. *Metaphysics of Meaning*. Cambridge, MA: MIT Press.

Kenny, A. 1975. *Will, Freedom and Power*. Oxford: Blackwell.

Kenny, A. 1989. *The Metaphysics of Mind*. Oxford: Oxford University Press.

Kenny, A. 2010. Concepts, Brains, and Behaviour. *Grazer Philosophische Studien* 81: 105-113.

Kim J. 2000. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim J. 2003. The American Origins of Philosophical Naturalism. In: R. Audi, ed. *Philosophy in America at the Turn of the Century*. *Journal of Philosophical Research*, APA Centennial Volume, p. 83-98.

Kim J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Kim J. 2007. Causation and Mental Causation. In: B. McLaughlin and J. Cohen, eds. *Contemporary Debates in Philosophy of Mind*. Oxford:

Blackwell Publishing, p. 227-42.

Kim, J. 1987. "Strong" and "Global" Supervenience Revisited. *Philosophy and Phenomenological Research* 48: 315-26.

Kim, J. 1990. Explanatory Exclusion and the Problem of Mental Causation. In: E. Villanueva, ed. *Information, Semantics, and Epistemology*. Cambridge, MA: Basil Blackwell, p. 36-56.

Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Lawrence, S. 2001. Naturalism and the Interpretation of Theories. *Proceedings and Addresses of the American Philosophical Association* 75(2): 43-58.

Levin, J. 2013. Functionalism. *Stanford Encyclopedia of Philosophy*, [online] available at: http://plato.stanford.edu/entries/functionalism/ [accessed 2015 10 27].

Lewis, D. 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy* 50: 249-258.

Lewis, D. 1980. Mad Pain and Martian Pain. In: N. Block, ed. *Readings in Philosophy of Psychology Volume I*. Cambridge, MA: Harvard University Press, p. 216-232.

Loar, B. 1987. Subjective Intentionality. *Philosophical Topics* 15(1): 89-124.

Loewer, B. 1987. From Information to Intentionality. *Synthese* 70: 287-317.

Loewer, B. 1997. A Guide to Naturalizing Semantics. In: C. Wright and B. Hale, eds. *A Companion to the Philosophy of Language*. Oxford: Blackwell: 108-126.

Lowe, E. J. 2007. *Personal Agency*. Oxford: Oxford University Press.

Luntley, M. 1999. *Contemporary Philosophy of Thought*. Oxford: Blackwell.

Lycan, W. 2003. The Mind-Body Problem. In: S. Stich and T. Warfield, eds. *The Blackwell Guide to Philosophy of Mind*. Oxford: Blackwell, p. 47-64.

MacDonald, C. 1989. *Mind-Body Identity Theories*. London: Routledge.

Marcus, E. 2009. Why There Are No Token States. *Journal of Philosophical Research* 34: 215-241.

Margolis, E., Laurence, S. 2007. The Ontology of Concepts – Abstract Objects

or Mental Representations? *Noûs* 41(4): 561-93.

McDowell, J. 1996. *Mind and World*. Cambridge: Harvard University Press.

McDowell, J. 1998a. *Meaning, Knowledge, and Reality*. Cambridge: Harvard University Press.

McDowell, J. 1998b. *Mind, Value, and Reality*. Cambridge: Harvard University Press.

McGinn, C. 1983. *The Subjective View*. Oxford: Clarendon Press.

McGinn, C. 1989. *Mental Content*. Oxford: Blackwell.

McGinn, C. 2002. *Knowledge and Reality: Selected Essays*. Oxford: Oxford University Press.

Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, R. G. 1995. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Mingers, J. 2001. Embodying Information Systems: the Contribution of Phenomenology. *Information and Organization* 11: 103-128.

Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Nimtz, C. 2009. Conceptual Truth Defended. In: N. Kompa, C. Nimtz, and C. Suhm, eds. *The A Priori and its Role in Philosophy*. Paderborn: Mentis, p. 137-155.

Nuccetelli, S. ed. 2003. Introduction: In: *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press, p. 1-22.

O'Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press.

Papineau, D. 1987. *Reality and Representation*. Oxford: Blackwell Publishers.

Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell Publishers.

Pavilionis, R. 1981. *Kalba. Logika. Filosofija: šiuolaikinių loginių-filosofinių kalbos koncepcijų analizė*. Vilnius: Mintis.

Pavilionis, R. 2005. *Prasmė ir tapatumas, arba Kelionė į save.* Vilnius: Mokslotyros institutas.

Peacocke, C. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.

Pickles, D. 1994. Reviewed Work: The Representational Theory of Mind: An Introduction. By Kim Sterelny. *The Philosophical Quarterly* 44(175): 252-254.

Pitt, D., 2012. Mental Representation. *Stanford Encyclopedia of Philosophy*, [online] available at: http://plato.stanford.edu/entries/mental-representation/ [accessed 2014 11 20].

Place, U. T. 1956. Is Consciousness a Brain Process? *British Journal of Psychology* 47(1): 44-50.

Plato. 2014. *Theaetetus*, trans. by J. McDowell. New York: Oxford University Press.

Price, H. H. 1953. *Thinking and Experience*. London: Hutchinson.

Putnam, H. 1967. Psychological Predicates. In: W. H. Capitan and D. D. Merrill, eds. *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, p. 37-48.

Putnam, H. 1973. Meaning and Reference. *The Journal of Philosophy* 70(19): 699-711.

Putnam, H. 1975a. The Meaning of "Meaning". *Minnesota Studies in the Philosophy of Science* 7: 131-183.

Putnam, H. 1975b. *Mind, Language and Reality: Philosophical Papers, Vol. 2*. Cambridge: Cambridge University Press.

Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Putnam, H. 1986. Meaning Holism. In: J. Conant, ed. *Realism with a Human Face*. Harvard: Harvard University Press, p. 278-303.

Putnam, H. 1988. *Representation and Reality*. Cambridge, MA: MIT Press.

Putnam, H. 1992. Response to John McDowell. In C. Hill, ed. *The Philosophy of Hilary Putnam: Philosophical Topics* 20(1): 347-408.

Putnam, H. 2004. The Content and Appeal of "Naturalism". In: M. De Caro and D. Macarthur, eds. *Naturalism in Question*. Cambridge, MA: Harvard

University Press, p. 59-70.

Quine, W. V. 1966. *Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press.

Quine, W. V. 1980. *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Quine, W. V. 1981. *Theories and Things*. Cambridge, MA: Harvard University Press.

Ramsey, P. F. 1931. *The Foundations of Mathematics and other Logical Essays.* London: Routledge & Kegan Paul.

Rosen, G. 2010. Metaphysical Dependence: Grounding and Reduction. In: B. Hale and A. Hoffmann, eds. *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press, p. 109-136.

Roughley, N. 2004. Naturalism and Expressivism: on the "Natural" Stuff of Moral Normativity and Problems with its "Naturalisation". In: P. Schaber, ed. *Normativity and Naturalism*. Frankfurt: Transaction Books, p. 47-86.

Rowlands, M. 1995. Externalism and Token-Token Identity. *Philosophia* 24(3): 359-375.

Rupert, R. 2008. Causal Theories of Mental Content. *Philosophy Compass* 3/2: 353-380.

Russell, B. 1905. On Denoting. *Mind* 14: 479-493.

Russell, B. 1912. *The Problems of Philosophy*. Oxford: Oxford University Press.

Russell, B. 1966. On the Nature of Truth and Falsehood. In: *Philosophical Essays*. London: Allen & Unwin, p. 147-159.

Ryle, G. 1971. *Collected Essays, Vol. II.* London: Hutchinson.

Ryle, G. 1974. Mowgli in Babel. *Philosophy* 49(187): 5-11.

Sandis, C. ed. 2009. Introduction. In: *New Essays on the Explanation of Action*. London: Palgrave Macmillan, p. 1-12.

Sawyer, S. 2011. Internalism and Externalism in Mind. In: J. Garvey, ed. *The

Continuum Companion to Philosophy of Mind. London: Continuum Press, p. 133-150.

Schantz, R. ed. 2004. Introduction. In: *The Externalist Challenge*. New York: Walter de Gruyter, p. 1-36.

Schiffer, S. 1972. *Meaning*. Oxford: Oxford University Press.

Schiffer, S. 2005. *The Things We Mean*. Oxford: Oxford University Press.

Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Sellars, W. 1963. *Science, Perception and Reality*. New York: The Humanities Press.

Shea, N. 2013. Naturalising Representational Content. *Philosophy Compass* 8(5): 496-509.

Sidelle, A. 1989. *Necessity, Essence, and Individuation: A Defence of Conventionalism*. Ithaca: Cornell University Press.

Slaney, K. L., Racine, T. P. 2011. On the Ambiguity of Concept Use in Psychology: Is the Concept "Concept" a Useful Concept? *Journal of Theoretical and Philosophical Psychology* 31(2): 73-89.

Smart, J. J. C. 1959. Sensations and Brain Processes. *Philosophical Review* 68: 141-156.

Soames, S. 2010. *What is Meaning?* Princeton, NJ: Princeton University Press.

Speaks, J. 2003. Three Views of Language & the Mind. *PhD Thesis*, [online] available at: https://www3.nd.edu/~jspeaks/papers/3views.pdf [accessed 2015 04 23].

Speaks, J. 2006. Is Mental Content Prior to Linguistic Meaning? *Noûs* 40(3): 428-457.

Stalnaker, R. 1984. *Inquiry*. Cambridge, MA: MIT Press.

Stalnaker, R. 1997. Reference and Necessity. In: B. Hale and C. Wright, eds. *A Companion to the Philosophy of Language*. Oxford: Basil Blackwell, p. 534-585.

Stalnaker, R. 1998a. What Might Nonconceptual Content Be? In: E.

Villanueva, ed. *Philosophical Issues 9*. Atascadero, CA: Ridgeview, p. 339-352.

Stalnaker, R. 1998b. Replies to Comments: [Neale, Acero, Pineda]. In: E. Villanueva, ed. *Philosophical Issues 9*. Atascadero, CA: Ridgeview, p. 389-395.

Stalnaker, R. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stalnaker, R. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

Stampe, D. 1979. Toward a Causal Theory of Linguistic Representation. In: P. French, T. Uehling, and H. Wettstein, eds. *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesota Press, p. 81-102.

Stanley, J. 2010. "Assertion" and Intentionality. *Philosophical Studies* 151(1): 87-113.

Steward, H. 1997. *The Ontology of Mind: Events, Processes, and States*. Oxford: Clarendon.

Steward, H. 2009. Sub-intentional Actions and the Over-mentalization of Agency. In: C. Sandis, ed. *New Essays on the Explanation of Action*. London: Palgrave Macmillan, p. 295-312.

Steward, H. 2013. Processes, Continuants, and Individuals. *Mind* 122(487): 781-811.

Stich S. P., Laurence, S. 1994. Intentionality and Naturalism. *Midwest Studies in Philosophy* 19(1): 159-182.

Stout, R. 2006. *The Inner Life of a Rational Agent: In Defence of Philosophical Behaviourism*. Edinburgh: Edinburgh University Press.

Stoutland, F. 2009. Determinism, Intentional Action, Bodily Movements. In: C. Sandis, ed. *New Essays on the Explanation of Action*. London: Palgrave Macmillan, p. 313-337.

Strawson, P. 1974. Causation in Perception. In his *Freedom and Resentment*

and Other Essays. London: Methuen, p. 73-93.

Strawson, P. 1979. Perception and its Objects. In: G. F. Macdonald, ed. *Perception and Identity*. London: Macmillan, p. 41-60.

Strawson, P. 1986. *Skepticism and Naturalism: Some Varieties*. New York: Columbia University Press.

Strawson, P. 1992. *Analysis and Metaphysics.* Oxford: Oxford University Press.

Strawson, P. 1998. Reply to Susan Haack. In: L. E. Hahn, ed. *The Philosophy of P. F. Strawson*. Chicago & La Salle: Open Court, p. 64-67.

Stroud, B. 2004. The Charm of Naturalism. In: M. De Caro and D. Macarthur, eds. *Naturalism in Question*. Cambridge, MA: Harvard University Press, p. 21-35.

Thomasson, A. 2007. *Ordinary Objects*. New York: Oxford University Press.

Thornton, T. 1998. *Wittgenstein on Language and Thought*. Edinburgh: Edinburgh University Press.

Timmons, M. 1999. *Morality Without Foundations: A Defense of Ethical Contextualism*. New York: Oxford University Press.

Tye, M., 1994. Naturalism and the Problem of Intentionality. In: P. French, T. Uehling, and H. Wettstein, eds. *Midwest Studies in Philosophy 19: Philosophical Naturalism*. Notre Dame: University of Notre Dame Press, p. 122-142.

von Wright, G. H. 1963. *Norm and Action: A Logical Enquiry*. London: Routledge & Kegan Paul.

Watson, J. B. 1913. Psychology as the Behaviorist Views it. *Psychological Review* 20: 158-177.

White, A. 1972. Mind-Brain Analogies. *Canadian Journal of Philosophy* 1(4): 457-472.

Wiggins, D. 2001: *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Wilson, J. 2014. No Work for a Theory of Grounding. *Inquiry* 57(5-6): 1-45.

Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

Wittgenstein, L. 1958. *The Blue and Brown Books*. Oxford: Blackwell.

Wittgenstein, L. 1967. *Philosophical Investigations*, trans. by G. E. M. Anscombe. Oxford: Blackwell.